AFRL-HE-WP-TR-1999-0215

UNITED STATES AIR FORCE RESEARCH LABORATORY

MODELS OF PILOT PERFORMANCE FOR SYSTEMS AND MISSION EVALUATION-PSYCHOLOGICAL AND PSYCHOPHYSIOLOGICAL ASPECTS

Erland Svensson
Maud Angelborg-Thanderz

NATIONAL DEFENCE RESEARCH ESTABLISHMENT
HUMAN FACTORS DIVISION
LINKOPING, SWEDEN

FOA

GLENN F. WILSON

HUMAN EFFECTIVENESS DIRECTORATE
CREW SYSTEM INTERFACE DIVISION
WRIGHT-PATTERSON AFB OH 45433-7022

20000320 016

AUGUST 1999

INTERIM REPORT FOR THE PERIOD MAY 1997 TO MARCH 1999

Human Effectiveness Directorate
Crew System Interface Division
2255 H Street
Wright-Patterson AFB OH 45433-7022

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from Air Force Research Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center
8725 John J. Kingman Road, Suite 0944
Ft. Belvoir, Virginia 22060-6218

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-WP-TR-1999-0215

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Instruction 40-402.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

MARIS M. VIKMANIS, DR-IV
Chief, Crew System Interface Division
Human Effectiveness Directorate
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
August 1999

3. REPORT TYPE AND DATES COVERED
Interim Report May 1997-March 1999

4. TITLE AND SUBTITLE

Models of Pilot Performance for Systems and Mission Evaluation - Psychological and Psychophysiological Aspects

5. FUNDING NUMBERS

C F41624-94-D-6000
PE 62202F
PR 7184
TA 14
WU 25

6. AUTHOR(S)

Erland Svensson*, Maud Angelborg-Thanderz* and Glenn F. Wilson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

*National Defence Research Establishment, Linkoping, Sweden

8. PERFORMING ORGANIZATION

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Human Effectiveness Directorate
Crew System Interface Division
Air Force Materiel Command
Wright-Patterson AFB, OH 45433-7022

10. SPONSORING/MONITORING

AFRL-HE-WP-TR-1999-0215

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release: distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The purpose of this study was to analyze the effects of mission complexity and information load on Pilot Mental Workload (PMWL), Situational Cognizance (SC), and Operative Effectiveness (OE) or operative performance. Previous assessment of operative performance or effectiveness has been developed with task analyses using psychological indices constructed with factor analysis. Generally, PMWL is affected by mission complexity and PMWL affects different aspects of Pilot Performance (PP). The specific purposes were to (a) validate psychological, psychophysiological, and performance based measures of PMWL, SC, and OE, (b) develop models of pilot performance for systems and mission evaluation, (c) compare real and simulated missions, and (d) discuss the application of these results to the systematic evaluation of systems and missions with the pilot in the loop. In the first phase, 20 fighter pilots performed 150 flight missions. In the second phase, 15 pilots performed 40 simulated missions. Questionnaires were used to tap mission complexity, information load, mental workload, mental capacity, motivation, situational cognizance, and performance. Additionally, during the simulated missions eye movements, heart and blink rates were obtained. From the flight and simulation data, the model analyses showed that mission complexity affects different aspects of information processing and mental workload and that these aspects, in their turn, affect situational cognizance and pilot performance. Significant relationships were found between heart rate, workload ratings, mental capacity, operative performance, and motivation. Heart rate and eye fixation rates increased, and blink rate decreased with increasing task complexity. A combination of these dynamic measures and the psychological indices form a reliable and valid technique for systems and mission evaluation.

14. SUBJECT TERMS
Mental Workload, Pilot Performance, Operator Modeling, Heart Rate, Eyeblinks

15. NUMBER OF PAGES
57

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Modern flight systems are complex, and they are getting even more complex. This greatly increases the amount of information the human operator has to process. As pilots have to work in a more complex situation today than before, their information load must be decreased, but current decision support systems do not meet such demands. During complex mission phases military pilots are reaching and passing capacity limits of human information processing. To analyze the actual cognitive needs of the pilots, models of pilot performance must be developed as well as reliable and valid methods to assess workload, situational awareness, and performance.

Specific purposes of the present study were to (a) validate psychological, psychophysiological, and performance based measures of pilot mental workload (PMWL), situational cognizance (SC), and operative effectiveness (OE), (b) develop models of pilot performance for systems and mission evaluation, (c) compare real and simulated missions, and (d) discuss the application of these results to the systematic evaluation of systems and missions with the pilot in the loop.

Decisions and Complexity of Information

Decisions are affected by uncertainty, ambiguity, and limited human capacity. It is hardly possible for anyone to be observant of everything at the same time. Even if military pilots are rigorously selected and trained, the corresponding demands on them to react rapidly and accurately in a synthetic and "hyperdynamic" environment often are extreme. For this reason, it is important to take into account what is known about the limitations of human information processing in research on pilot mental workload (PMWL) and operative effectiveness (OE).

When too many alternatives with too many attributes compete for attention, the decision maker will be mentally overloaded. Mental overload often has its origin in cognitive limitations. Miller (1956, cp., Baddeley, 1994) found that humans cannot discriminate between more than half a dozen one-dimensional entities. Nor can they handle more objects in their short-term memory or control more content in their attention.

Many people feel safer when they have a great deal of information, even if they do not use it. There is an illusion of knowledge (van Raaij, 1988). The quality of the decisions decreases with increasing amount of information beyond the optimum, at the same time as the decision maker's illusion increases.

The time factor is in itself a stress factor and, of course, important in an analysis of the pilot's judgments and decisions in a rapidly changing environment. Generally, psychological stress can induce tunnel vision and a more primitive motor behavior (Easterbrook, 1959). It can also cause an emphasis on negative and threatening information (Svenson and Edland, 1987). On

the other hand it is not certain that availability of more time will lead to better decisions (Svenson and Benson, 1991).

Those pilots who can best integrate "performance critical" information have a Position Of Advantage (POA, cf., Jane's Aerospace Dictionary, 1986). This ability forms the core of the concept of Situational Awareness (SA) for which the skill aspect is of importance. Even if the theoretical anchorage of SA still is weak, it can be assumed that it calls for automatic information processing and an efficient memory function (cf., Gilson, 1995).

Klein (1993), in his naturalistic decision making models, regards the pilot's ability to recognize and assess the situation (situation assessment) as the crucial factor of the decision making process. For the skilled pilot almost every situation is associated with a "best alternative."

The limitations on human information processing have been known for a long time. In spite of this, designers are apparently tempted by the possibilities, created by modern computer technology, to include increasingly complex and numerous options (modes) and displays in their systems. The result is that human operators are faced with very complex tasks which tax their mental capacities.

Pilot Mental Workload (PMWL), Its Rationale and Measurement

Pilots in modern flight systems (military as well as civilian) have to process a considerable amount of complex information, much more comprehensive than in older systems. The information and decision making processes have become more and more demanding and the risk for mental overload has increased (Angelborg-Thanderz, 1990; Kantowitz, 1993; Svensson et al., 1997). Procedures for measuring mental workload have become an indispensable prerequisite for analyzing specific flight tasks, evaluating the need of decision support systems (automation, data fusion, artificial intelligence, and expert systems), and in evaluating effects of training procedures. Efficient training is reflected in better performance, lower workload, and increased reserve capacity (Svensson et al., 1993).

During the last two decades great effort has been spent on defining and measuring the concept of mental workload (Williges and Wierwille, 1979; Moray, 1979; Hancock and Meshkati, 1988; Lysaght et al., 1989; Hart and Wickens, 1990; Eggemeier, Wilson, 1991; Farmer, 1993; Carmody, 1994). One general conclusion is that PMWL is a multifaceted concept, hard to define and measure. It cannot be measured validly and reliably by one single measure (Gopher and Donchin, 1986).

Two different elements recur in attempts to define workload: (a) What is required by the pilot? What is the pilot expected to accomplish with his flight system? (b) Under which conditions must the flight be performed? For (a) it is assumed that PMWL increases as a function of the number of tasks and the difficulty of the separate tasks. For (b) it is assumed that PMWL in-

creases as a function of the number and significance of unfavorable conditions which can be external (e.g., extreme temperatures) or internal (e.g., psychological stress, mental fatigue, insufficient experience and training) and impair the pilots' capacity to process information, make decisions and act (Gawron, Schflett and Miller, 1989).

The relation between PMWL and PP is affected by the pilots' capacity to cope with the demands of the flight tasks. Gopher and Donchin (1986) define PMWL as "the difference between the capacities of the information processing system that are required for task performance to satisfy performance expectations and the capacity available at any given time" (p. 41-3). Hart and Wickens (1990) define PMWL "as the effort invested by the human operator into task performance" (p. 258).

As can be seen from the above discussion, there does seem to be a fair amount of agreement among current experts that mental workload is a multidimensional concept involving an interaction between pilot, task, and environment (Carmody, 1994).

Main PMWL Measurement Techniques In Use

The techniques for measuring PMWL can be divided into three main categories: subjective measures, performance or task measures, and psychophysiological measures. A further category includes analytical measures, which examine the relation between required and available time to perform tasks. The analytical methods (time-line analysis) are of special importance in the initial phase of the systems design process.

The subjective rating methods require the pilot to rate his mental workload directly on a specific scale or that he rates different aspects of the concept. In the latter case, the aspects can be merged (sometimes by differential weights) into a workload index. The procedure assumes that the pilot can accurately rate his workload or its manifestations.

Task measures require that the pilot's performance is measured either on a primary task (e.g., maintaining assigned altitude) or on a secondary task (e.g., reacting to warning devices). It is assumed that decreased performance indicates increased workload (Eggemeier and Wilson, 1991).

The use of psychophysiological measures presumes that physiological reactions are related to the demands of the task. The psychophysiological reactions can be mediated by emotional stress, an increased psychological activation, preparedness, and effort (Wierwille, 1979). Heart rate, heart rate variability, event-related potentials, eye blink activity, pupillary dilation, and endocrine reactivity are examples of different psychophysiological techniques (Wilson and Eggemeier, 1991).

Subjective measures

Subjective ratings seem to be the most utilized technique, sometimes used in combination with a physiological measure (especially heart rate) (Roscoe, 1987; Roscoe and Ellis, 1990; Eggemeier and Wilson, 1991). Casali and Wierwille (1983) recommend subjective techniques because they are cheap, easily administrated, and adaptable to different situations. Modified versions of the Cooper-Harper Handling Characteristic Scale (Cooper and Harper, 1969; Wierwille and Casali, 1983; Wierwille and Connor, 1983; Roscoe, 1987) have frequently been used in the aircraft industry to measure pilot workload.

The BedFord Rating Scale (BFRS) is a decision tree scale derived from the Cooper-Harper Scale (Roscoe, 1987; Roscoe and Ellis, 1990). According to Lysaght et al., (1989), "The technique obtains subjective judgements about workload based on ability to complete tasks and the amount of spare capacity available" (p. 88). The "Continuous Subjective Assessment of Workload" (C-SAW) technique (Jensen, 1993) is based upon the Bedford scale.

The Subjective Workload Assessment Technique (SWAT) (Reid and Nygren, 1988) and the NASA Task Load indeX (NASA-TLX) (Hart and Staveland, 1988) exemplify frequently used multidimensional rating scales. In SWAT the aspects of time load, mental effort load, and stress load are merged into a workload index by means of conjoint measurement. NASA-TLX measures the aspects of mental, physical, and temporal demands, as well as performance, effort, and frustration levels. The aspects are differentially weighted and merged into a single workload index.

The global measure such as NASA-TLX, SWAT, and MCH (Modified Cooper-Harper) are about the same with respect to sensitivity and diagnosticity. However, NASA-TLX and SWAT provide some diagnosticity through the different dimensions of which they are composed. Especially SWAT and MHC have very high acceptance by the aviation community (Eggemeier, Biers, Wickens, Andre, Vreuls, Billman and Schueren, 1990). The same is true for BFRS. MCH and BFRS are very easy to implement.

Many researchers consider that subjective techniques have high face validity. Johannsen, Moray, Pew, Rasmussen, Sanders, and Wickens (1977) claim that "Despite all the well-known difficulties of the use of rating scales we feel that these must be regarded as central to any investigation. If the person feels loaded and effortful, he is loaded and effortful whatever the behavioral and performance measures may show."

As compared to performance based measures, subjective ratings are sensitive to ranges of workload below the point of overload. It is sometimes claimed that the reliability and validity of subjective ratings of PMWL and PP are insufficient (Muckler and Seven, 1992), and that it can be difficult to fully ascertain what has been measured. All mental processes are not introspectively available. Accordingly, the subjective measure can yield an underestimation of workload. Retrospective ratings are affected by memory deficiencies (Gopher and Donchin,

1986; O'Donnell and Eggemeier, 1986). It has been found that the workload is higher, when it is rated close to the peak workload (Carmody, 1994). The ratings can also be affected by the answering patterns of the participants.

Subjective measures have, however, turned out to be extremely useful in many contexts. Doubts about their validity, while sometimes justified should not be exaggerated. Even if the precision of any single rating is modest, data may still be sufficiently rich in information to be useful. Reliabilities of the mood scales used in our research have been found to range between 0.65 and 0.95 (Sjöberg, Svensson and Persson, 1979). The reliability (Cronbach's alpha) of ten psychological indices used in a simulation study ranged from .67 to .90 (Svensson, Angelborg-Thanderz, Sjöberg and Olsson, 1997).

Performance measures

Performance measures of PMWL are objective, non-intrusive, and have high pilot acceptance. However, the pilots´ ability to compensate for increased task demands renders performance measures less useful as measures of PMWL. They are most sensitive, when the demand of the situation exceeds the pilots´capability to process information. In system development and evaluation it is of importance to get information about increases in PMWL before it is manifest in decreased performance.

Another objection to the use of performance measures of PMWL is that performance measures are important in their own right and it is confusing to use them as aspects of PMWL.

Psychophysiological measures

Heart Rate (HR)

The general assumption behind the use of different psychophysiological measures of PMWL is that a pilot's physiological activation level is affected, when his/her mental capacity is challenged (Carmody, 1994). Heart rate (HR) has, since the twenties, been the most popular physiological variable to monitor the state of human operators during flight, and it has been the most used psychophysiological measure of PMWL (Caldwell, Wilson, Cetinguc, Gaillard, Gundel, Lagarde, Makeig, Myhre and Wright, 1994; Carmody, 1994). The interactions between the sympathetic and the parasympathetic nervous systems influence HR. Both systems are affected by higher cortical centers.

A wealth of empirical data shows that this measure has sensitivity under different circumstances in both simulated and real flights (Eggemeier, at al., 1990; Wilson and Fullenkamp, 1991). HR has been demonstrated to covary with workload associated with different mission or flight segments in a variety of aircraft systems. The measure has been found to differentiate between crew-members, and it has been used under most realistic situations (e.

g., short landings on Swedish road bases) (cf., Angelborg-Thanderz, 1982; Wilson, Skelly and Purvis, 1988; Wilson, 1991). The present authors have used the technique (HR) in combination with measures of endocrine reactivity (adrenaline and noradrenaline) and subjective measures in both applied and simulated missions. Angelborg-Thanderz (1990) reports significant correlations between HR and adrenaline reactivity ($r = .81$, $p < .01$) and HRV and adrenaline reactivity ($r = .68$, $p < .01$) over flight missions. HR in combination with subjective measures has often been used in the aircraft industry to measure PMWL (Roscoe, 1987; Roscoe and Ellis, 1990; Roscoe, 1992).

It has been noted that HR sometimes fails to distinguish between variations in task load when subjective measures have been found sensitive. The tasks manipulated in such cases have been perceptual load and central processing load during simulated flights (Eggemeier, et, al., 1990; Casali and Wierwille, 1983). Differences as well as structural similarities between simulated and real flight have been found in several studies (Angelborg-Thanderz, 1990; Wilson, Purvis, Skelly, Fullenkamp and Davis, 1987; Wilson, 1991, 1993). Even if the changes in HR are lower during simulated as compared to real situations, there is a covariation between the simulated version and the real version of a specific flight or mission (Angelborg-Thanderz, 1990).

Unlike the usual subjective PMWL measures, heart rate is registered continuously (is a dynamic measure). This makes an analysis of changes in psychophysiological activation as a function of changes in, e.g., information load during flights possible. In a study of PMWL and PP, we have found that HR (running means) covaried significantly with variations in information load over the missions for those pilots who performed above mean performance. The correlation between the pilots' rank order with respect to PP and the covariations (HR - information load) was -0.55 ($p = .019$). Thus, the sensitivity and diagnostic value of HR were affected by the pilots' skill level. Figure 1 presents the covariation between HR (running means) and the information load (number of objects) presented on Tactical Situation Display (TSD) as a function of mission time for an "expert" pilot. The common variance between the curves is 55 percent. Thus, about half of the variance of HR is explained by the variance in information load (Svensson et al., 1997).

That HR can be a sensitive and reliable dynamic measure of PMWL has been shown in a series of studies. Systematic relations between cognitive demands and HR have been found in simulated as well as real situations (Wilson and Eggemeier, 1991; Caldwell et al., 1994). For example, a test-retest correlation of .67 ($p < .01$) for peak HR indicates an acceptable reliability (Angelborg-Thanderz, 1990). However, physical activity should be monitored and taken into consideration when interpreting results.

Heart rate has high applicability. As noted above, HR can be and has been used in a broad range of tasks, manipulations, and environments. However, movements, muscle activity, and Respiration Rate (RR) can contaminate the measure.

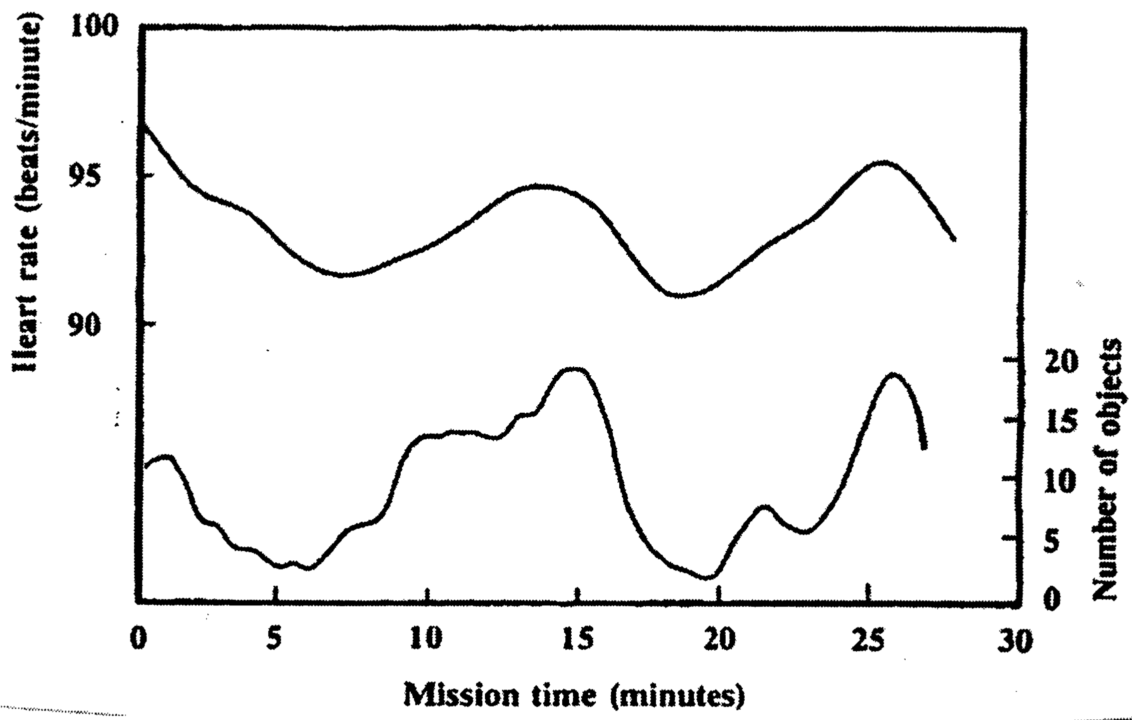

Figure 1. Heart rate (beats per minute) (upper curve) and number of objects on TSD (Tactical Situation Display) (lower curve) as a function of mission time for an experienced pilot. (From Svensson et al., 1997). The curves have been smoothed by means of distance weighted least squares regression.

Eye blink activity

Visual information gathering and visual attention play dominant roles in most cockpit tasks. Eye blink activity has been and is used as an indicator of PMWL in situations where high visual attention is required. Eye blink data has been collected in highly realistic settings. Blink Rate (BR), Blink Duration (BD), and Blink Latency (BL) have been analyzed and used as PMWL measures in a series of studies (Eggemeier et al., 1990; Wilson and Fisher, 1991; Wilson, 1993). Both BR and BD decrease with increases in task demands. The latency measure has been found to increase with memory demands (Eggemeier, et al., 1990). Blink rate has been found to be sensitive and capable to differentiate among mission types (Wilson, O'Donnell and Wilson, 1982), and it has been found to decrease significantly during high load segments of missions (Wilson and Fisher, 1991). Blink patterns can be used to provide information about pilots' response to different stimuli and thus situational awareness. A general conclusion is that BR may be most related to visual information requirements. Blink duration and BL show more promise as measures of PMWL (Carmody, 1994). Wilson and Fisher (1991) have demonstrated the advantage to use both HR and eye blink data in analyses of PMWL.

Eye blink is a dynamic measure sensitive to high visual attentional demands and to pilot fatigue (Stern, Boyer, Schroeder, Touchstone and Stoliarov, 1994). Fewer and shorter duration blinks are related to situations that require intake of important information (e.g., flying) (Wilson et al., 1987). According to Wilson and Eggemeier (1991) additional studies need to be performed to properly assess the reliability of eye blink measures in operational situations.

Changes in eye blink activity can be caused by many factors. Meaning and diagnosticity must be attached to the eye blink activity by theoretically and empirically linking it to other variables or factors. We consider eye blink activity (BD, BR, and BL) to represent one facet of PMWL, and it is the co-variation of this facet with others that gives the construct PMWL its meaning.

Measures of eye blink activity (EOG measures) are non-intrusive and have high pilot acceptance, if adequate electrodes are used. Eye blink measures do not intrude on the pilots' primary tasks and is not a flight safety risk. Eye blink activity can also be measured by means of equipment for EPOG measurement (cf.,VINTHEC technical report II, 1997).

Eye blink measures have high applicability. As noted above, eye blinks can be and have been used in a broad range of tasks, manipulations, and environments. The measures have high face validity in terms of behavior and performance.

Eye point of gaze (EPOG)

Modern military cockpits provide the pilots with a large amount of synthetic information about aircraft internal system functions, weapon systems functions, and the external combat environment. Vision plays a dominant role in most cockpit tasks and the pilots' visual information gathering performance is reflected and limited by eye movements (Katoh, Kadoo, Itoh and Maruta, 1995; VINTHEC-WP3-TR01, 1997). Scanning behavior and fixation times are related to different aspects of PMWL (Harris and Christhilf, 1980; May, Kennedy, Williams, Dunlap and Brannan, 1990; Itoh, Hayashi, Tsukui and Saito, 1990; Kennedy, Braun and Massey, 1995; Svensson, et al., 1997).

In Svensson et al. (1997) it was found (a) that the frequencies of shorter fixation times HU (Head Up) *and* the frequencies of longer fixation times HD (Head Down) increased as a function of the information load on TSD (Tactical Situation Display). Thus, the condition for flying low level-high speed with high precision deteriorated when the information load HD increased. It was also found that the durations and frequencies of critical eye fixations HD (fixations equal to or longer than 4 seconds) covaried with PMWL. The correlation between the frequency of critical fixations HD and ratings of PMWL on the Bedford scale was .51 ($p < 0.01$).

Situational Awareness (SA)

The SA concept continues to have a major impact on the aviation research community, despite the fact that there is no agreed upon definition (McMillan et al., 1995). In fact at least 15 more or less different definitions have been formulated. According to Endsley (1995) Situational Awareness (SA) can be described as a person's state of knowledge or mental model of the situation around him. A general applicable definition describes SA as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (Endsley, 1988). The operational community of USAF has adopted the following definition of SA: "A pilot's continuous perception of self and aircraft in relation to the dynamic environment of flight, threats, and mission and the ability to forecast, then execute tasks based on that perception." The definition was developed by an Air Staff SA working group (Carrol, 1992; McMillan et al, 1995). As noted, the concept has its origin in the aircraft domain, and most definitions of the concept are based on experiences from this area.

The concept is considered to be of importance in the evaluation of new flight and weapon systems and in the evaluation of pilot training. Explicit measures of SA during system development and testing can help to determine to what extent the design objectives have been met. Problems with SA brought on by, e.g., high information load and high information and system complexity can be detected, if direct measures of SA are performed.

It is important to establish the relationships between SA and PMWL and PP. According to Endsley (1995) SA is considered to be a precursor to the pilot's decision making and a stage separate from his performance. According to our opinion and experience, SA is often so closely related to the pilot's performance that it is more logical to consider it a part of performance. In our research, operative pilot performance includes SA aspects such as target detection and target identification (Angelborg-Thanderz, 1990). In Svensson et al (1992, 1997) a significant correlation ($r = .59$, $p < .001$) was found between Tactical Situational Awareness (TSA) and flight performance. Those pilots who performed well with respect to the flight task also performed well with respect to the information handling task and vice versa. Experienced pilots have developed internal mental models of the systems they operate and the environments they operate in. The pilots have learned to assess the situation by means of critical cues. In Klein's model of Dynamic Decision Making (DDM) (Klein, 1993) situation assessment is the central aspect of the pilot's decision making. Thus, SA is dependent on pilots' ability to match patterns between critical cues in the environment and their elements of his mental model.

The relationship between SA and PMWL has been considered to be of theoretical importance. However, according to Endsley (1995), SA and PMWL vary independently over a large range of the spectrum. Only when the workload demands exceed the mental capacity, the SA is at risk. Of course, the pilot's SA can be reduced, when the workload is very low due to vigilance problems.

Numerous techniques for measuring SA have been proposed. The range of the spectrum is from verbal protocols and direct subjective questions (e.g., what is your SA just now?) to direct objective experimental techniques (e.g., different probe techniques or the registration of the pilot's Eye Point Of Gaze (EPOG) on the panels and on instruments in the cockpit). The most common approach of the experimental or objective techniques is SAGAT (Situation Awareness Global Assessment Technique) developed by Endsley (1995).

In VINTHEC (Visual Interaction and Human Effectiveness in the Cockpit) a technique for subjective assessment of SA has been developed. The scale is derived from the BFRS (the Bedford rating scale of mental workload) and the decision tree structure with three main levels, a satisfactory, an acceptable but not satisfactory, and an unacceptable level of SA has been utilized. It has been used and analyzed by FOA at TFHS (the Swedish commercial flight school). The pilots answered the scale after each of seven segments of two flight scenarios. We found that the scale was functional and it could be used repeatedly with minimum interference. The SA-ratings were significantly related to both mental workload and performance. But the relation between the instructors ratings of the pilots SA and the pilots own ratings was low (cf., Berggren, 1998).

Operational Performance Criteria

In the procedure of defining criteria of the very complex behavior of a modern pilot, there are some obstacles to be overcome. There are skills, which are tacit, hidden, and imbedded. The pilot's handling of the system may have a wide range of effects from immediate to gradual or remote and from trivial to critical. The outcome may be manifest by very simple actions or no action at all.

The four "fundamental problems" that Vreuls and Obermayor (1986) identify in a prize article in Human Factors are a long way from being solved. (The article discusses simulator performance measurement. According to our experience the conclusions drawn are equally valid for measurement in the air, in some sense even more, as there are other practical difficulties to be added.) The four problems are briefly described below.

1. *Hidden knowledge and embedded performance* as mentioned above.

2. *Lack of theories of performance* means that too many investigators are driven to collect "a large amount of useless data for a given task and environment"... "In the absence of theories to guide selection of performance measures, one is driven to the alternative of measuring as much as reasonably possible" (p. 243).

3. Studies of *measurement validity* are usually lacking which is connected with the fact that researchers seldom know enough of...

4. *...operational performance criteria.* "Researchers seldom know the operational meaning of a performance change in their experiment" (p. 242). Vreuls and Obermayor think it is rare to find other criteria differentiating between novices and experts than a number of something, e.g., flight hours. They conclude that "these metrics are useful to describe experience, but they are not performance criteria" (p. 245).

Soon enough we realized that we must focus on actions of decisive importance for the outcome, events that differed a success from a failure. The so-called objective criteria were not enough due to the problem of hidden knowledge, embedded performance, and to the fact that the outcome could be manifest by very simple actions or no action at all. Questions to and answers from observers and participants are even more important, as "almost all important behavior are cognitive... Since all the essential behaviors take place "in the head," objective measurement of operator behavior is insufficient" (Meister and Hogg, 1994).

Often we must rely on questions to and answers from both the instructor and the pilot himself. Our procedure was to make detailed check-list-like questionnaires tailored for different intercepts. Depending on all of the circumstances, the number of questions could differ resulting in questionnaires of different size. Too many details led to a loss of lucidity and precision. The border between too many and too few items could vacillate depending on the task.

We were fortunate to get assistance from very skilled expert pilots. Our method hinged on the participation of these experts. When the instructors developed an insight in their students' reasoning, they saw that they could more efficiently correct the students. This led to a complete analysis of a fighter pilot's job in two steps: analysis of the situations and analysis of the pilots' actions within the situations.

Research on expertise is most intriguing in this context (cf., e.g., Chi, Glaser and Farr (1988). Learned skill and factual knowledge, including knowledge of strategies, seem to be the dominant source of performance differences between experts and novices (Simon, 1990).

Thanks to the fact that our research from the Swedish Air Force's point of view has been decision-oriented and furthermore done in a close dialogue with pilots - who know the operational meaning of their performance - we have often enough been able to validate our simulator results in the air (cf., Thanderz, 1973, 1982; Angelborg-Thanderz, 1989, 1990; Svensson and Angelborg-Thanderz, 1994, 1995; Svensson, Angelborg-Thanderz and Sjöberg, 1991, 1993a, 1993b; Svensson, Angelborg-Thanderz, Sjöberg and Gillberg, 1988). From the end of the seventies those validation studies have included psychophysiological variables.

To summarize, intercepts were chosen which constitute a concrete objective blueprint of the training.

Then we have split the intercepts into smaller parts, pilot actions which are logical with regard to enemy threat and factual weapon system. We have defined questions concerning the pilot's

behavior on the basis of the skills and the performances we hope to achieve in our pilots (and which we have found in skilled pilots). We have been very careful taking - and defining - only those actions of decisive importance to an outcome. It has sometimes been enough to have the instructor - or the pilot - answer "yes" or "no," but sometimes it was more appropriate to have them evaluate behavior in more detail. That is whether a "bad" action, had it a large influence on the outcome of the mission? We have often used a five-point scale from "no influence at all" to "crucial influence." We made detailed questionnaires tailored to the different intercepts.

After simulator training, the instructors have usually answered the questions. After intercepts in the air, the pilot himself has answered in the presence of an experienced instructor.

The present study represents a step in a series of studies with the purpose to analyze the effects of mission complexity and information load on Pilot Mental Work-load (PMWL), Situational Cognizance (SC), and Operative Effectiveness (OE). Specific purposes of the study are to (a) validate psychological, psychophysiological, and performance based measures of PMWL, SA, and OE, (b) develop models of pilot performance for system and mission evaluation, (c) compare real and simulated missions, and (d) discuss the application of these results to the systematic evaluation of systems and missions with the pilot in the loop.

MISSIONS IN THE AIR

Method

Two regular training periods at Blekinge Air Force Base (Wing F17) during autumn and early winter 1995 were judged as appropriate for the study. Two squadrons participated during the first period. One squadron acted as a fighter interceptor group (FIG) and the other acted as the enemy in a ground attack. The enemy came in a column at low level and were escorted by fighters. During the second period the enemy aircraft, now without fighter escort, manoeuvred to escape the fighters. The Swedish multi-role aircraft 'Viggen' was the system used in the study.

Scenarios

Air tasking order codes. (101) Air defence VMC. The mission was to fight an escorted attack unit at low level. No Swedish fighters were specifically allocated to engage hostile escorting fighters. (102) Air defence VMC. The mission was to fight an escorted attack unit at low level. Some Swedish fighters were specifically allocated to engage hostile escorting fighters. (103) Air defence IMC. The mission was to fight attack units in bad weather. (104) Air defence in darkness. (105) Aggressor mission. This code represented all the various hostile missions. (106) Air defence VMC. The mission was to fight a low-level attack unit without escort. When engaged by

Swedish fighters, the attack aircraft manoeuvred trying to escape. (100) Special exercises with non-regular pilots.

The mission types 102, 105, and 106 represent each about 25 percent of the missions performed. Mission type 104 represents 12 percent, and the types 100, 101, and 103, 13 percent.

Subjects

Twenty active fighter pilots from two squadrons at Wing F17 participated and performed 150 missions of which 144 have been used in the analyses. The pilots' mean time in combat aircraft was 395 hours (standard deviation = 137 hours).

Measures Used

The pilots answered check lists before and after each mission. Before the missions they rated their motivation, expected performance, perceived mental and physical state, and mood in terms of the following dimensions: hedonic tone, activation, tension, and control (Sjöberg et al. 1979, Svensson et al. 1988, 1993, 1997).[1]

After the missions the pilot answered 90 assorted items, mission mood, and six items of NASATLX, BFRS, and final mood. The pilots answered these items under supervision; back checking did not occur.

The post-mission questionnaire tapped aspects of mission difficulty, perceived performance, motivation, control, vigilance, mental capacity, mental and physical effort, situational cognizance, concentration, information load (TSD; Tactical Situation Display, and TI; Target Indicator), priority to tasks, interference between tasks, availability, and complexity of information (TSD and TI). Most of the items have been used in former studies (cf., Svensson et al., 1997).

By means of iterative principal factor analyses, the number of items was reduced to 10 indices. The reliability of the indices was tested by means of Cronbach's alpha. The six aspects of NASATLX as well as the items of the questionnaire employed a seven-point response format. The items of NASA TLX were equally weighted.

The JA37 system is equipped with a registration system (UTB) that records mission and flight data. A ground based system is used to replay the missions. This equipment is a powerful training tool used by the pilots in de-briefings. It is also a powerful tool for analyses of pilot performance. Factors behind the pilots' performance can be scrutinized by the pilots and their

[1] The pre-mission ratings will be analyzed and presented in a master's thesis in psychology at the university of Uppsala by captain Johan Ginyard, Swedish Air Force Reserve.

instructors using detailed questionnaires tailored for different missions. The intention in this study was to record and analyze each mission by means of this technique in order to get an optimal estimate of the pilots' operative performance with respect to reliability and validity. Unfortunately, this technique of performance estimation often is time consuming, and in this study a tight training schedule hindered us. We were not able to analyze more than about 15 percent of the missions performed.

Statistics

Primarily, correlation statistics (simple and multiple regression, principal- and maximum likelihood factor analysis, and second generation multivariate statistics) were used. Linear causal model analyses were performed by means of LISREL VI (Jöreskog and Sörbom, 1984). This procedure makes a statistical test of the validity of different causal flow models.

Results

The post-mission questions formed the basis for analyses and diagnoses of mental workload, situational cognizance, and operative performance. By means of iterative principal axis factor analysis, the large number of items was reduced to 9 indices. The number of factors or indices was determined by means of the scree-test. This procedure provides a solution with the minimum number of factors accounting for the maximum amount of variance. The number of markers of the indices varies between three and seven, which means that the indices reflect different aspects of the concept. This multifacetedness supports both the reliability and validity of the concepts.

The reliabilities of the indices were estimated by means of Cronbach's alpha. Table 1 presents the indices and the reliability values. The indices are in the range of acceptable to high reliability. Perceived performance (PP) had the lowest reliability (.74), which means that the common variance between the markers of the index is 55 percent. Most of the indices have been found in former studies and, accordingly, this study validates these concepts (cf., Svensson et al., 1997).

The Situational Cognizance index illustrates how this latent variable is manifested in seven variables: To what extent could you estimate the flight paths (angle of advance), was the course of event as expected, could you predict the mission course of events, was the cooperation within the group functioning well? Did you have 'mental lead' with respect to the air defence task, did you recognize the course of events, and were you in control of the situation? It is interesting to note that this empirical factor or index reflects the important aspects of the definitions of both the US Air Force (McMillan, 1994) and Endsley (1995).

Table 1. Nine indices of analysis and their reliability.

Index	Chronbach's alfa
Percieved Performance (PeP)	.74
Situational Cognizance (SC)	.80
Difficulty (DIFFIC)	.84
Mental Effort (EFF)	.86
Pilot Mental Workload (PMWL)	.87
Ment. Capacity Red. (CAPAC)	.77
Motivation (MOTIV)	.84
Comp.Inform. TSD (COMP TSD)	.92
Comp.Inform. TI (COMP TI)	.93

In order to get a 'first opinion' about if and how the central concepts PMWL, SC, and PP change as a function of the complexity of the missions, we divided the missions into five groups. Group A consists of quite simple training missions and group E of applied missions of very high complexity. The groupings were based on complexity estimations made by the pilots during the briefings. Figure 2 presents the changes of means of PMWL, SC, and PP as a function of increases in mission complexity. From analyses of variance we found significant changes over the groups for the three indices.

First, an almost linear increase in mental workload over the five groups can be seen from the figure. At the workload level of group C it can be seen that the pilots mental reserve capacity is restricted. The workload means of the groups D and E indicate that the pilots try to shut themselves off from information, because they must focus on those aspects they consider most important. The expression 'Mental tunnel vision' can summarize this condition.

It can also be seen that situational cognizance and performance decrease over the groups. To begin with, the decreases are small but the changes of groups D and E indicate critical decreases in situational cognizance and performance.

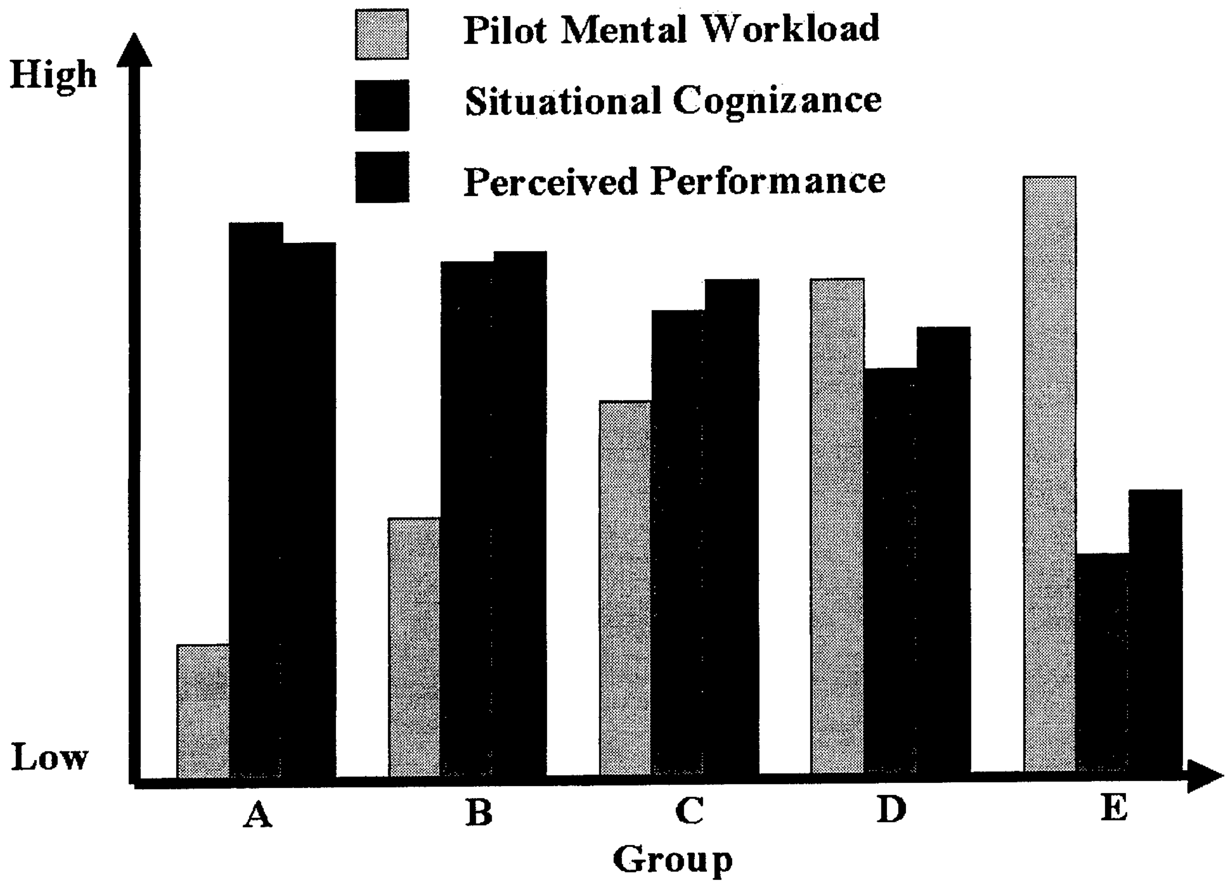

Figure 2. Changes in Pilot Mental Workload (PMWL), Situational Cognizance (SC), and Perceived Performance (PP) as a function of mission complexity. Group A consists of quite simple missions and group E of missions of very high complexity. Each group represents about 30 missions.

We have previously found that PMWL is comparatively sensitive to increased information load. Increases in workload turn up earlier than, and they can therefore predict, later decreases in situational awareness and performance (Svensson, 1997).

The conclusions based on the relative changes of the indices found in Figure 2 and results of former studies form an embryo to a model about how the concepts of mission complexity or difficulty, pilot mental workload, situational cognizance, and performance are related to and affect each other. Figure 3 illustrates the proposed model.

In order to test the credibility of the causal model presented above we have used structural equation modeling ad modum LISREL (Jöreskog and Sörbom, 1984). By means of this technique we can test the statistical goodness of fit of a specific model in the population from which the sample has been drawn.

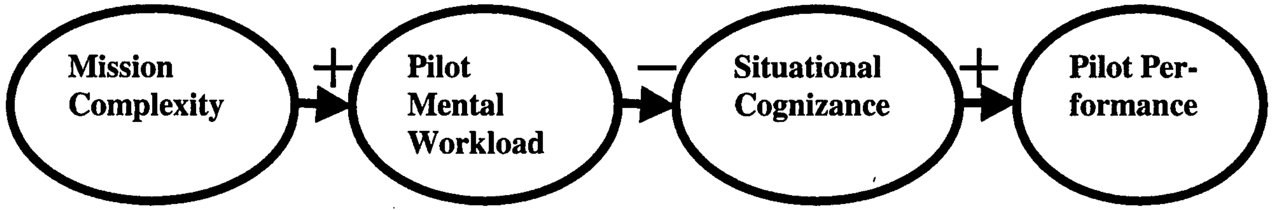

Figure 3. A model of causal relationships between the concepts mission difficulty, pilot mental workload, situational cognizance, and pilot performance.

In the analyses we have used the following factors or indices from Table 1: mission difficulty (DIFFIC), complexity information Tactical Situation Display (COMP TSD), complexity information Target Indicator (COMP TI), mental capacity reduction (CAPAC), situational cognizance (SC), and perceived performance (PeP). Pilot mental workload was measured by means of BFRS (the Bedford rating scale of mental workload; (Roscoe, 1987; Roscoe and Ellis, 1990). The final model is presented in Figure 4. The model analysis is based on the correlations (product moment) between the markers of the indices.

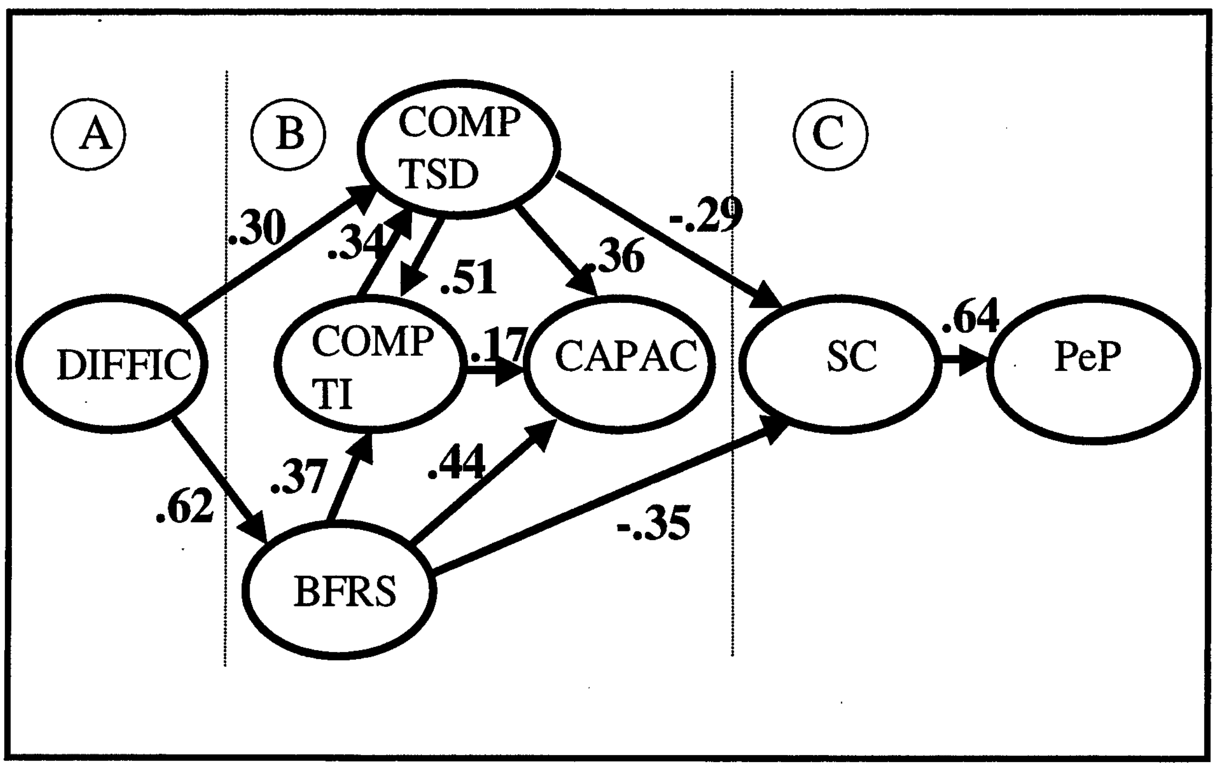

Figure 4. The final structural LISREL model of the relationships between six of the indices and the BFRS workload scale. All effects are significant ($p < .05$). Adjusted Goodness of Fit is .85 and Root Mean Square is .053.

The circles represent the different indices or factors and the arrows the directions of the effects. A positive sign means that an increase in one index gives an increase in another and a negative sign that an increase in one factor gives a decrease in another. The effects can be considered as regression or normalized beta weights ranging from -1.00 to 1.00.

The fit of the model is high (Adjusted Goodness of Fit Index =.95 and Root Mean Square=.053). This means that the model can be generalized to the pilot population of the system. The Root Mean Square index presents the mean differences between the correlations of the input matrix and the corresponding correlations of a matrix reconstructed from the model. Thus, the mean difference between correlations is not more than about five hundredth (5/100).

As can be seen from Figure 4, the model has its starting point in the difficulty and complexity of the missions and its terminal point in the performance of the pilot. Increasing mission difficulty is followed by an increased general mental workload (BFRS) but, furthermore, the complexity of the synthetic information on the Tactical Situation Display (TSD) and the Target Indicator (TI) increases. That the information on these two displays is closely related, is supported by the fact that the fit of the model is highest when these two indices are interacting (indicated by two contrary directed arrows in Figure 4). The correlation between the indices COMP TSD and COMP TI is .77, which means that the common variance is 59 percent. Figure 5 presents the regression of general workload (BFRS) and information complexity (COMP TSD) on mission difficulty (DIFFIC). As can be seen, the increase in difficulty has stronger and earlier affects on general workload than on the information complexity of TSD. The fact that general workload affects the information complexity on the interacting TSD and TI in the model is supported by the differences between the curves of Figure 5.

From the model we find that increases in general mental workload (BFRS), in their turn, reduce mental capacity (CAPAC). Increases in information complexity on TSD and TI also yield a reduction of mental capacity of about the same size. Regression analyses show that the common effect of COMP TSD, COMP TI, and BFRS accounts for 65 percent of the variance of the mental capacity index. The six markers of the mental capacity index (CAPAC) deal with difficulties concerning evaluations of synthetic information and the necessity or need to reduce the flow of information. In other words, the model tells us that there is a strong connection between the information load on the displays and mental overload or a reduced mental reserve capacity.

It is also evident from the model that increases in general workload (BFRS) and information complexity on TSD (COMP TSD) both decrease situational cognizance of the pilots. That the pilots' situational cognizance grows worse as a function of high information complexity is a memento. The anticipated effect of the mental capacity index on situational cognizance was not found.

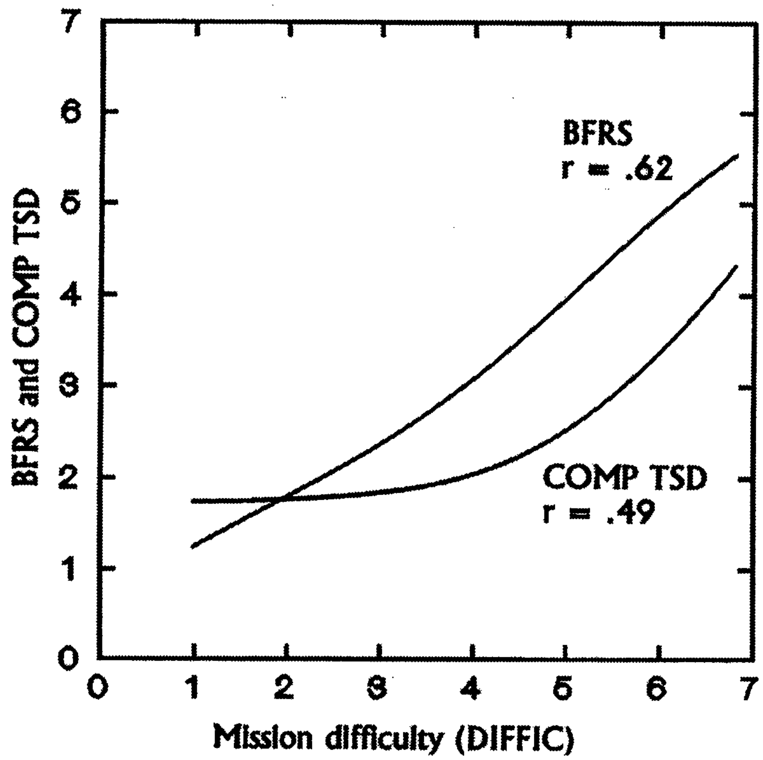

Figure 5. The increase in mental workload (BFRS) and information complexity on TSD (COMP TSD) as a function of mission difficulty (DIFFIC). The curves have been smoothed by means of distance weighted least squares regression analyses.

The relationships between the markers of the indices COMP TSD and COMP TI have been analyzed by means of multidimensional scaling (MDS; Schiffrin, Reynolds and Young, 1981). This procedure fits our variables in an Euclidean space in such a way that the distances between them corresponds to their inter-correlations. Figure 6 presents a two-dimensional MDS solution for the markers of COMP TSD and COMP TI.

From the figure it can be seen that dimension II separates the TSD items (squares) from the TI items[2] (circles). It can also be seen that dimension I arranges the markers in a sequence common to both indices (the same items of the two indices are connected with lines in the figure). The sequences are shown by the dashed arrows. (The fit of the data is almost perfect and the relations between the items could be described in terms of distances on a plane.)

When analyzing the sequences we found that the left ends represent items of perceptual content (e.g., difficulties in surveying the symbolic representations), and that the right ends represent

[2] The same items (except for changes of display names) were used as markers for COMP TSD and COMP TI, respectively.

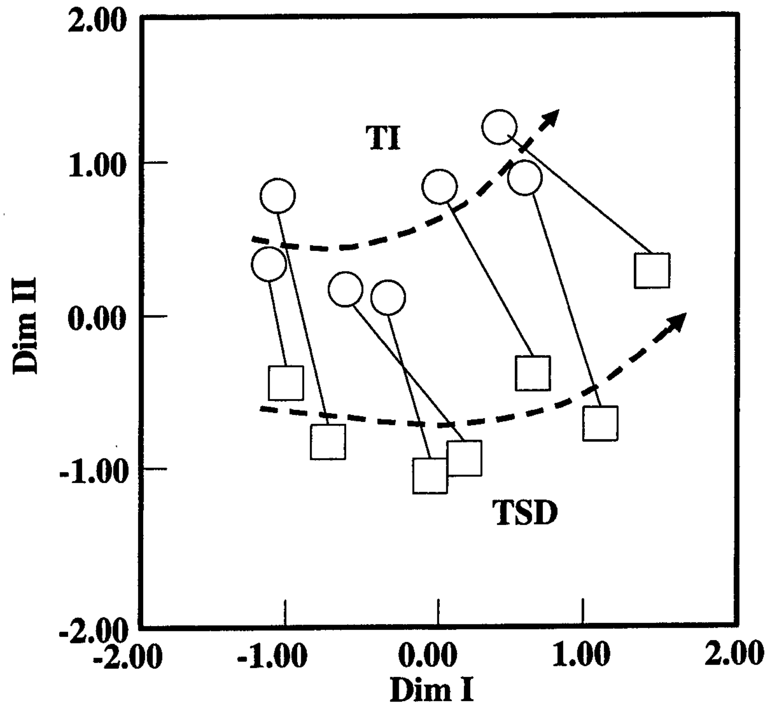

Figure 6. Two-dimensional MDS solution for the relations between the markers of COMP TSD and COMP TI. According to Guttman-Lingoe's coefficient of alienation, 99 % of the variance in the data is explained by the solution.

items of cognitive content (e.g., difficulties in integrating information and make decisions). These sequences are in agreement with the Radex theory. The Radex theory prescribes a different measurement model than the factor theory[3].

From the items representing the perceptual and cognitive ends respectively, we have formed four sub-indices; COGTSD, COGTI, PERCTSD, and PERCTI. Figure 7 presents a two-dimensional MDS solution of the relations between two of our workload measures (the workload index and the Bedford scale) and the four sub-indices.

As can be seen from Figure 7, the workload measures are more related to the cognitive aspects of COMP TSD and COMP TI than to the perceptual aspects. Table 2 presents the multiple regressions of the sub-indices on the workload measures. As can be seen from the table only the beta weights of the cognitive aspects are significant. About 20 to 30 % of the variance in mental workload is explained by the variance in the cognitive aspects.

[3] In factor theory the factor score is the 'sum of the markers' of a factor and the order of the summation is unimportant. In radex theory the order is important (e.g., variable B implies variable A; C implies A and B; and D implies A,B, and C).

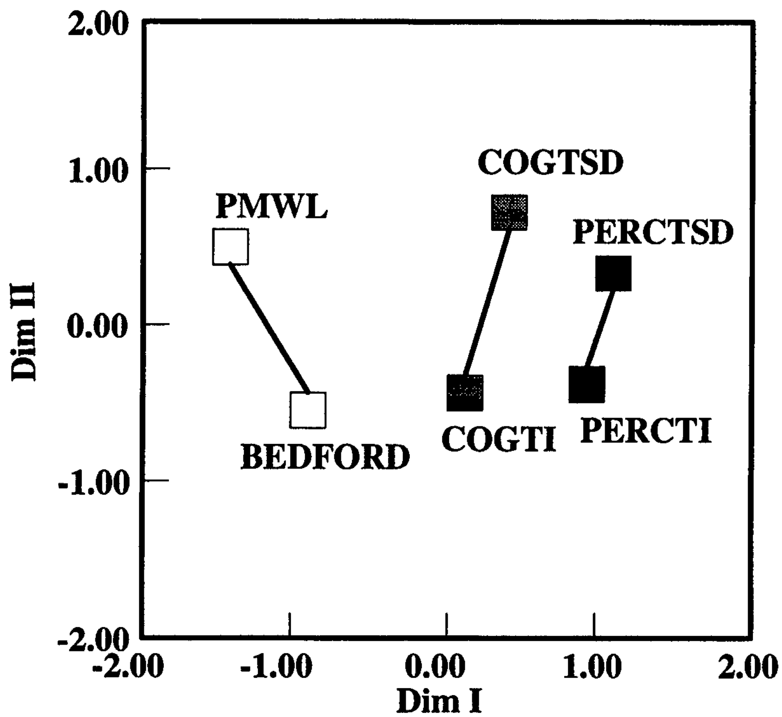

Figure 7. Two-dimensional MDS solution for the relations between the workload measures [the workload index (PMWL) and the Bedford scale (BEDFORD)] and the sub-indices COGTSD, COGTI, PERCTSD, and PERCTI. According to Guttman-Lingoe's coefficient of alienation, 99 % of the variance in the data is explained by the solution.

PMWL

In earlier studies we have found a close relationship between situational awareness and pilot performance and, as can be seen from the model (Figure 4), this was the case in this study too. The pilot's situational awareness is a predictor of his performance.

Table 2. Multiple regression analyses of effects of perceptual and cognitive aspects on mental workload. COGTI = cognitive aspects TI, PERCTI = perceptual aspects TI, COGTSD = cognitive aspects TSD, and PERCTSD = perceptual aspects TSD. Bold type beta weights are significant ($p < .01$).

PMWL = **.49 COGTI** - .06 PERCTI, $R^2 = .20$, ($F = 14.54$, df = 2; $p < .001$)

BFRS = **.37 COGTI** + .16 PERCTI $R^2 = .27$, ($F = 17.91$, df = 2, $p < 001$)

PMWL = **.52 COGTSD** - .04 PERCTSD, $R^2 = .24$, ($F = 18.48$, df = 2; $p < .001$)

BFRS = **.40 COGTSD** + .04 PERCTSD $R^2 = .19$, ($F = 11.52$, df = 2, $p < .001$)

Figure 8 presents the pilots' performance as a function of their situational cognizance. The relationship is moderate and the variance in situational cognizance explains 41 percent of the variance in performance. As can be seen from the figure, the curved regression deviates from the diagonal in the lower end, and low situational cognizance is related to higher levels of performance.

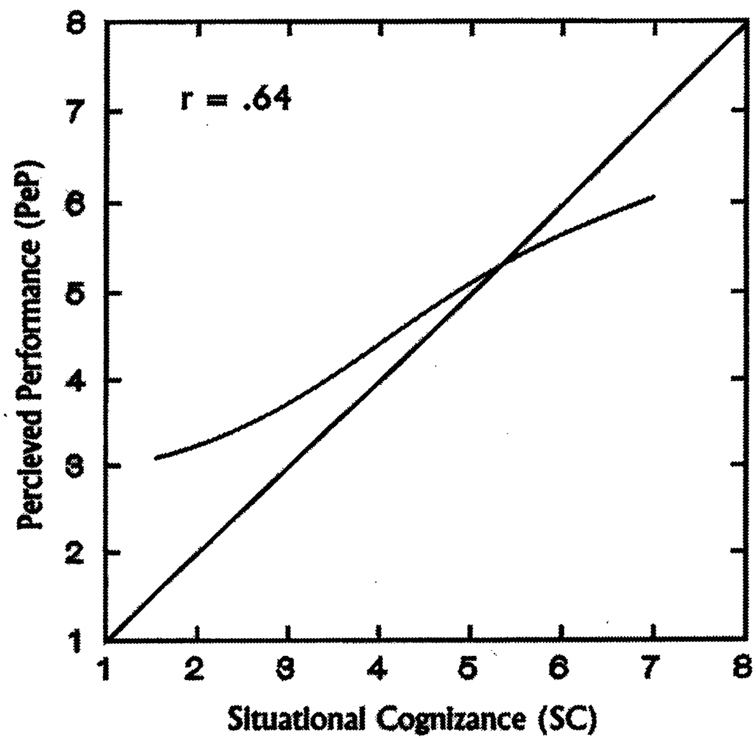

Figure 8. Pilot performance (PeP) as a function of situational cognizance (SC). The curves have been smoothed by means of distant weighted least squares regression analyses.

Figure 9 illustrates the direct and indirect causal flow effects of mission difficulty (DIFFIC) on the dependent indices of the model. As can be seen, the effects of mission difficulty (in terms of common variance) decrease as a function of the distances to the other indices. The effect of mission difficulty on performance is not more than 5 percent. On the other hand, the indices of the model that are more closely related to performance account for 40 percent of the variance.

The model (Figure 4) can be divided into three consecutive parts; A, B, and C. Part A consists of aspects of missions and systems demands, part B comprises aspects of mental workload,

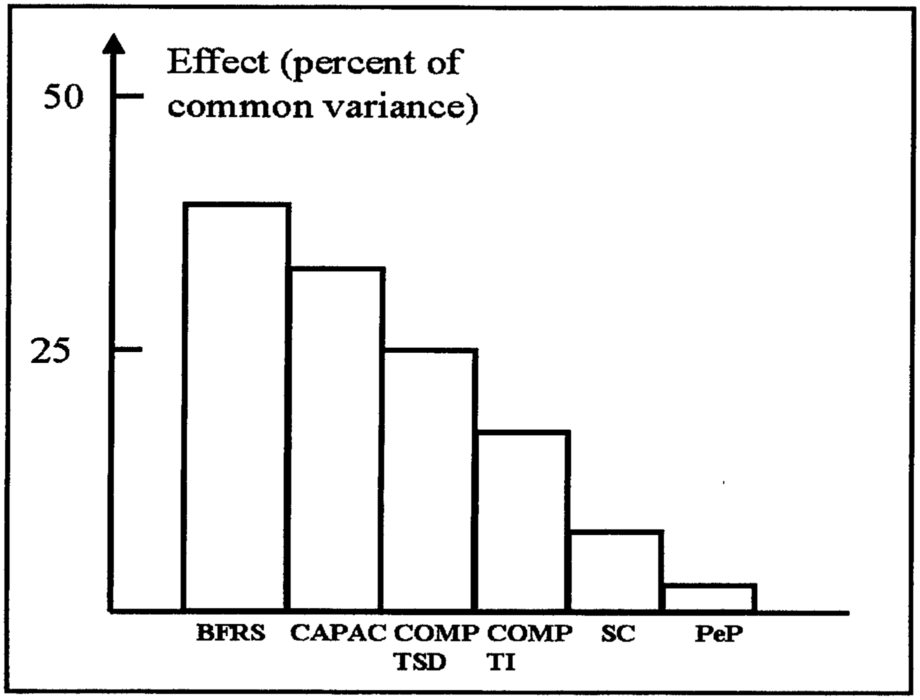

Figure 9. Direct and indirect effects of mission difficulty (DIFFIC) on the dependent indices of the model in terms of common variance.

information load and mental capacity, and part C includes situational cognizance and performance aspects. This means that there is a close correspondence between the proposed model in figure 3 and the final model in Figure 4. The causal sequence of systems and mission demands to mental workload to situational cognizance to pilot performance has come up in other studies. The way the pilot copes with the demands of the mission forms an intermediary and compensating process affecting his performance.

Discussion - Missions In the Air

Factor analyses of the data from missions in the air validated the psychological indices developed in former studies and we found that they have an acceptable to high reliability. Subjective ratings made by experts have turned out to be extremely useful in many contexts. The precision of single ratings is often modest, but the grouping of ratings of different aspects into factors increases both reliability and validity.

We found that pilot mental workload is comparatively sensitive to increased information load. Increases in workload turn up earlier than decreases in situational cognizance and performance. These differences reflect how the pilots cope with the load of the situation. The pilots try, as long as possible, to maintain their performance and situational cognizance by increasing their mental effort. However, the performance decrements show that this compensation only works so long.

The mental workload of high complexity missions indicates that the pilots mental reserve capacity is restricted. Pilots try to shut themselves off from information, because they must focus on those aspects they consider most important. The expression 'Mental tunnel vision' summarizes this condition.

We found an unfavorable relationship between mental workload, situational cognizance and performance during missions of moderate and high complexity. It is reasonable to suppose that the mission complexity and psychological stress of real war situations are higher than that of even the most complex missions of this study. Thus, extrapolation of the changes in mental workload, situational cognizance, and performance found in this study indicates a sub-optimal performance of the pilots in real war situations.

From model analyses we found that mission complexity affects different aspects of information and mental workload and that these aspects, in their turn, affect situational cognizance and pilot performance. The model tells us that there is a strong connection between the information load on the displays and a reduced mental reserve capacity or 'mental overload'. It is also evident from the model that increases in general workload and information complexity both decrease the situational cognizance of the pilots. The pilots' situational cognizance grew worse as a function of high information complexity on the displays indicating a 'bottle neck' in the system.

When analyzing, the internal structure of the indices 'complexity of information on tactical situation display (COMP TSD) and 'complexity of information on target indicator' (COMP TI), we found that the markers could be ordered in a sequence. One end represents items of perceptual content, e.g., difficulties in surveying the symbolic representations, and the other end represents items of cognitive content, e.g., difficulties in integrating information and make decisions. Thus, the perceptual and cognitive aspects of information load were identified by the indices. It is interesting to note that the workload measures are more related to the cognitive aspects of COMP TSD and COMP TI than to the perceptual aspects (only beta weights of the cognitive aspects are significant). These findings support the position that cognitive aspects of information handling play a dominant role in cockpits of modern aircraft.

MISSIONS IN THE SIMULATOR

Methods

Beginning in the seventies we have conducted research, initiated by the Swedish Air Force, using its flight training simulators as a research platform in analyses of pilot performance and in development of models of pilot performance (cf., Angelborg-Thanderz, 1990, Svensson et al., 1997). In the present study, the presentation system of the JA37-simulator at Wing 17 was modified and the same registration system (UTB) as the one used in the aircraft was installed.

Subjects

Fifteen fighter pilots from two squadrons at Blekinge Air Force Base (Wing F17) performed a total 40 simulated missions. Thirty-five have been used in the analyses. The pilots' mean time on combat aircraft was 420 hours (standard deviation = 121 hours). The study was carried out during August 1996 (period one) and during September 1997 (period two). During both phases the pilots performed two different types of missions.

Scenarios

In simulation period one, the observed pilot served as a flight leader and the simulator instructor as his wingman. Together, their aircraft formed a Swedish air defence fighter unit. Their task was to distract and lure away escorting enemy fighters from a large enemy attack column, so that other Swedish fighters could engage this attack column. The enemy attack column approached at low level with Mach 0.7-0.8 and took no evasive action. Some aircraft in the formation were specially equipped for ECM[4]. The escorting enemy fighters were positioned at tactical altitude close to the column front and rear, and were controlled by the simulator instructor.

The study contained two types of missions. During mission number one, the Swedish fighters were exposed to radar jamming by units in the attack column. Communication was not jammed. When the mission was completed, the Swedish fighters returned to a suitable base. During mission number two, the Swedish fighters were exposed to both radar and communication jamming. When the mission was completed, the simulation was stopped and all units were reset to their starting position and status.

During simulation period two, the subject was pilot in a solitary fighter. The enemy attack column approached at low level with Mach 0.7-0.8 and took no evasive action. Some aircraft in the formation were specially equipped for ECM. The column was divided into two parts by a larger spacing between aircraft in the middle. Period two also contained two types of missions, and in each sortie only one mission type was carried out. During each sortie, though, the mission was carried out twice. In between, the aircraft was reloaded in the air and the pilot was evaluated.

In mission number one, the Swedish fighter had full support from ground control. The spacing between the two column parts was approximately 20 kilometers. Within the column parts, the spacing between aircraft was 8-10 kilometers. The Swedish fighter was exposed to radar jamming by units in the attack column. When the mission had been carried out twice, the fighter was reset to its starting position and status. In mission number two, the Swedish fighter had no support from ground control and, hence, operated totally autonomously. The spacing between the two column parts was approximately 20 kilometers. Within the column parts the spacing between aircraft was 5-7 kilometers. The Swedish fighter was exposed to radar jamming by units

[4] Electronic countermeasures.

in the attack column. When the mission had been carried out twice, the fighter was reset to its starting position and status.

Our intention was to prepare scenarios of very high complexity with respect to the information load on the pilots without violating the realism. The intention was to a large extent fulfilled even if the simulator and its presentation system were confining.

Measures Used

Before and after each mission, the pilots answered the same questionnaires as in the study in the air and the same indices were used. (The scale format, the data reduction, and the reliability of the indices were presented in the section on missions in the air.)

After each of the two intercepts during simulation period two, the pilots responded to questions about mental workload (using the Bedford Scale BFRS) and performance on a seven point Likert scale. Situational awareness was assessed using a scale developed within the VINTHEC-project. (Svensson, Angelborg-Thanderz and van Avermaete, 1997).

During the simulated missions, the pilots Heart Rate (HR) and Blink Rate (BR) were recorded. A 'Del Mar Neurocorder' recorder was used during the first phase and a 'VitaportII' recorder was used during the second phase. Silver/silver chloride electrodes were placed on the sternum and intercostal space to record HR, and above and below one eye to record BR. An electrode on the right side of the chest served as ground.

The mean HR for a two-minute period was found centered around the maximum HR during each period of interest. The periods of most interest were the intercept phases. In simulation phase two, we also used the approach phases as comparisons to the intercept phases.

The pilots Eye-Point-Of-Gaze (EPOG) was video taped (Phillips CCD camera LDH0460/00). The durations and frequencies of eye fixations on seven different areas of the instrument panel were recorded manually from the video tape. Those areas were the Tactical Situation Display (TSD), the Target Indicator (TI), the Head Up Display (HUD), two side panels, and outside view left and right. In this study, we have used Fixation Rate (FR) as a general index of the pilots' visual search behavior. The measure indicates how often the eye fixation changes from one area to another per unit time (30 seconds). Analyses of where the pilot was looking and the sequence of his fixations are of great interest with respect to training and cockpit design and in the VINTHEC project we tried to find procedures to present and aggregate this type of information.

Specific purposes of the simulation studies were to validate psychophysiological measures of PMWL and to relate these measures to psychological measures of PMWL, SC and OE.

Statistics

The Heart Rate (HR) and Blink rate (BR) data were analyzed by means of the Workload Assessment Monitor (WAM). The EPOG-data were analyzed manually by means of a video recording system. We used the same statistical procedures as in the study of missions in the air.

Results

Except for the psychophysiological variables (HR, BR, and EPOG) the same measures were used in the simulator as in the air. One important advantage of the psychophysiological variables is that they are continuous and reflect the changes of a dynamic situation. As a first step we analyzed the 'mean two minutes HR' measure and related it to the psychological indices. A reliability analysis of the HR measure is presented first. A wealth of empirical data shows that Heart Rate (HR) is a sensitive measure under different circumstances in both simulated and real flights. Figure 10 presents an especially lucid example of how a pilot's HR can change as a function of the different phases of a mission. The figure illustrates different phases of interest on which we have calculated running maximum two-minutes HR means.

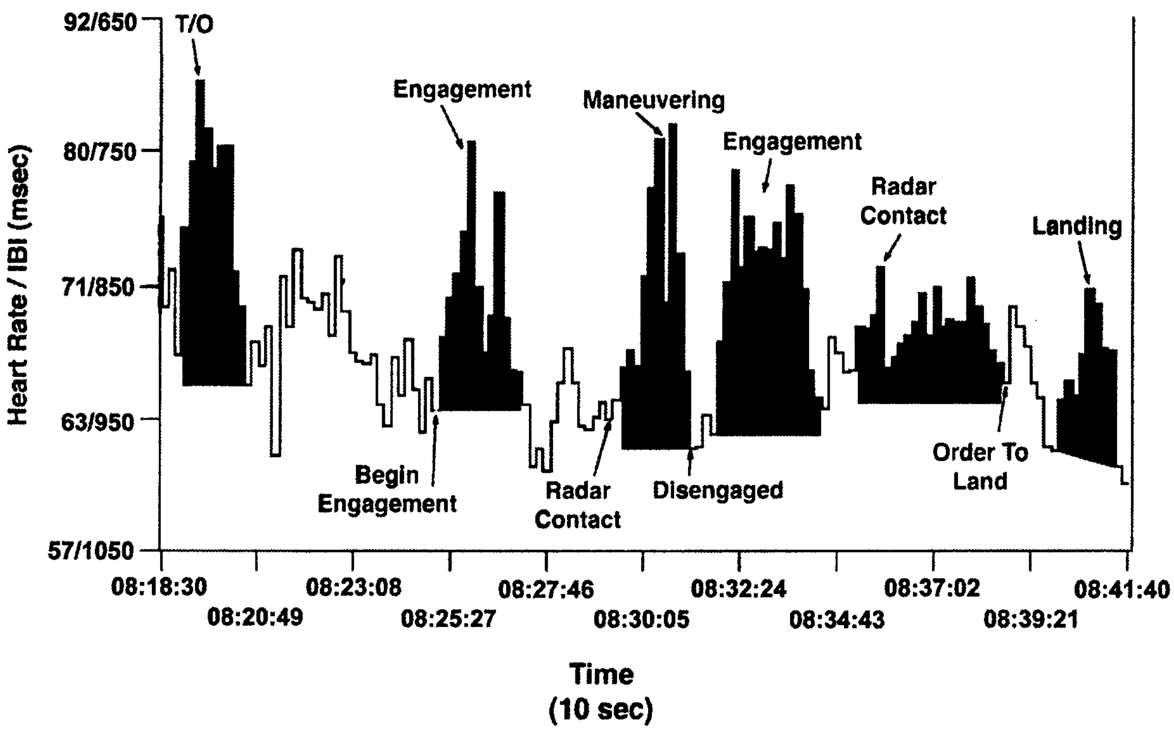

Figure 10. Illustration of how a pilot's HR changes as a function of the different phases of a mission. Each bar represents the HR mean for 10 seconds. Periods of interest in this example are black.

Even if the course of events of the scenarios is controlled, the timing and length of the different phases of a mission differ from trial to trial. This is a natural consequence of the situation and the interactions between the contracting parties. Unfortunately, it complicates our ability to aggregate the dynamic changes over subjects and missions. We have tried to find statistical procedures for this aggregation.

The missions of simulation phase two comprised two intercepts (intercept A and B) of comparable complexity. In order to test the sensitivity of the maximum two-minute HR means, we compared the means of intercept A with the corresponding means of a preceding approach period. The means of intercept B were compared with the means of a preceding low workload period during which the pilots answered questions about intercept A. Figure 11 presents the means of the four periods.

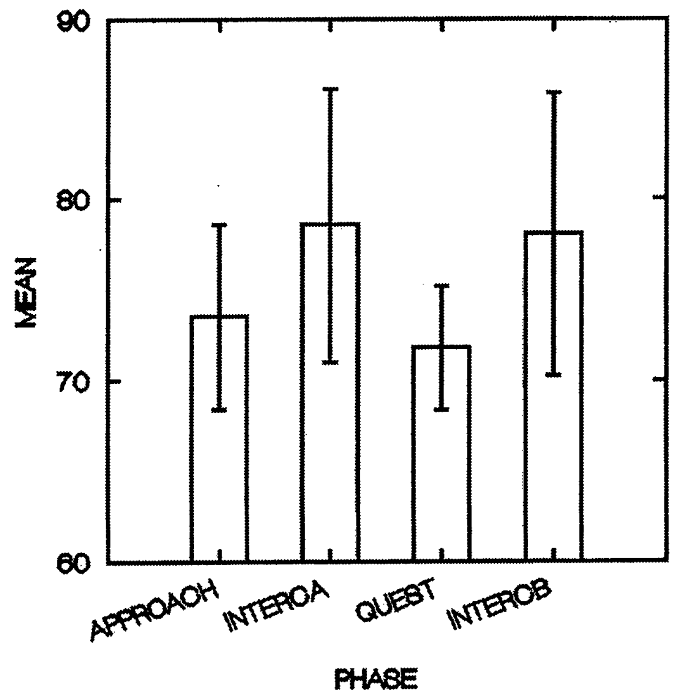

Figure 11. Mean HR (two minute means) and standard deviations for simulation phase two. The two minute means are shown for the approach (APPROACH), intercept A (INTERCA), question period (QUESTION), and intercept B (INTERCB).

A one-way ANOVA (F=3.887, p< 0.14) showed that the intercept heart rates were significantly higher than those during the approach and question periods. No significant differences were found between the two intercepts or between the approach and question periods.

We have used the correlation between the two intercept phases as an estimate of the reliability of the two minutes mean HR measure. The correlation is .82 ($p < .001$), meaning that 68 percent of the HR variance in the second intercept can be explained by the variance in the first.

The differences between the mean HR during the approach and intercept phases has been used as crude measure of the HR reactivity. In order to analyze the stability of this reactivity measure, we have compared the HR reactivity during intercept A and intercept B. A tendency to a co-variation was found ($r = .52$; $p < .06$), which indicates that there seems to be inter-individual differences in HR reactivity. The fact that the reactivity measures include two error terms instead of one deflates the magnitude of the correlation.

The fixation rate (FR) was analyzed in the same way as the HR-measure. When comparing the mean of the fixation rate for intercept A with that of intercept B we found a significant co-variation ($r = .73$; $p < .01$). This indicates that 53 percent of the variation in fixation rate of intercept B is explained by the variation in the fixation rate in intercept A. This shows that the FR measure is reliable and that there are inter-individual differences in the visual search behavior of the pilots. This stability of the visual search behavior of the pilots has implications for both training and cockpit design. To the present authors knowledge, this stability from intercept to intercept has not been shown and empirically validated before.

The same indices (see Table 1) were used during the simulated missions as in the missions in the air. Several significant correlations were found between the indices and mean HR. The HR measure correlated .67 ($p < .001$) with a reduced mental capacity (CAPAC), -.45 ($p < .032$) with Performance (PeP), .41 ($p < 054$) with Motivation (MOTIV), and tended to correlate negatively with SA ($r = -.38$; $p = .073$).

The correlation structures (i.e., the correlation matrices for the nine indices of Table 1) from the simulator and the air studies were compared. A positive correlation of .75 ($p < .001$) was found between the structures (see Figure 12). As can be seen in the figure, the overall relationship is close to the diagonal for the negative correlations but diverging from the diagonal for the positive correlations. Compared to the corrrelations from the air the positive correlations from the simulation are deflated.

The relationship between the motivation and difficulty indices was different and an obvious outlier, as seen in the lower right of the figure. Motivation was positively associated with difficulty in the air but negatively in the simulations.

The concordance between the two matrices means that the internal relationships between the indices are about the same in the data from the air and the data from the simulation. Accordingly, it is reasonable to use the model from the air (Figure 4) as a starting point in the

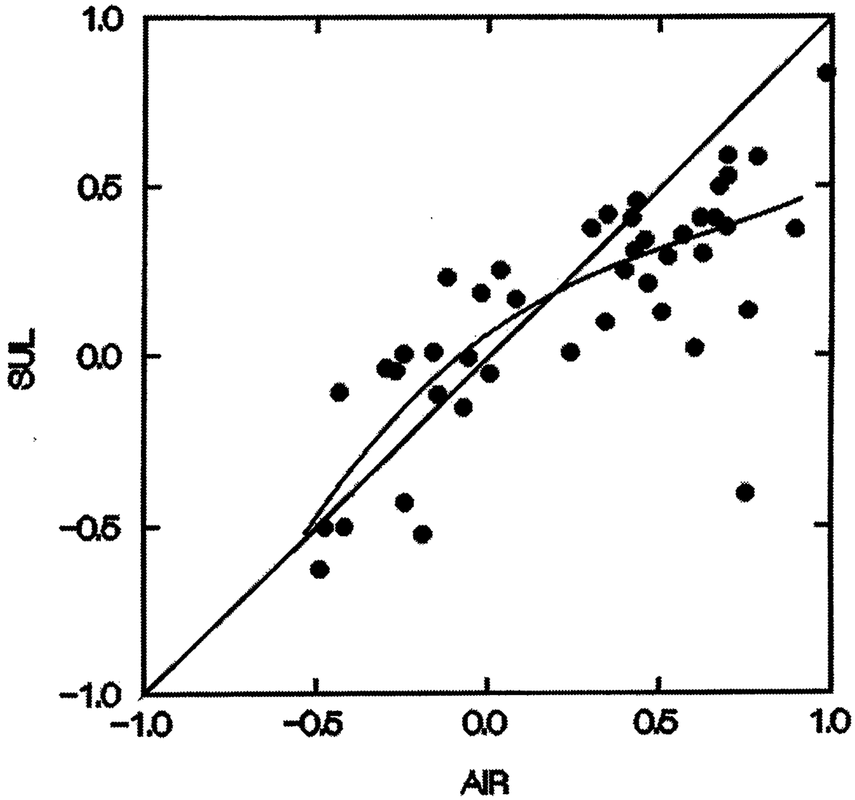

Figure 12. The relation between the correlations of the matrices from the air (AIR) and from the simulations (SUL). The correlation between the structures is .75 ($p < .001$).

model analyses of the data from the simulations. In order to permit the testing of a model representing the simulated missions, the data from phase one and phase two of the simulations were combined. To the psychological indices, the mean HR measure was added in the analyses. When missions contained two intercepts, a mean of the HR measures was derived.

In the analyses, we have used the following factors or indices from Table 1: mission difficulty (DIFFIC), complexity information Tactical Situation Display (COMP TSD), complexity information Target Indicator (COMP TI), pilot mental workload (PMWL), mental capacity reduction (CAPAC), situational cognizance (SC), perceived performance (PeP), and mean two-minutes HR (HR). The final model is presented in Figure 13. The model analysis is based on the correlations (product moment) between the markers of the indices.

The fit of the model is modest (Adjusted Goodness of Fit Index =.62 and Root Mean Square=.119). The fit of this model is lower than that of the model from the air. One statistical reason is the low number of cases (35) used to derive the correlations used in the

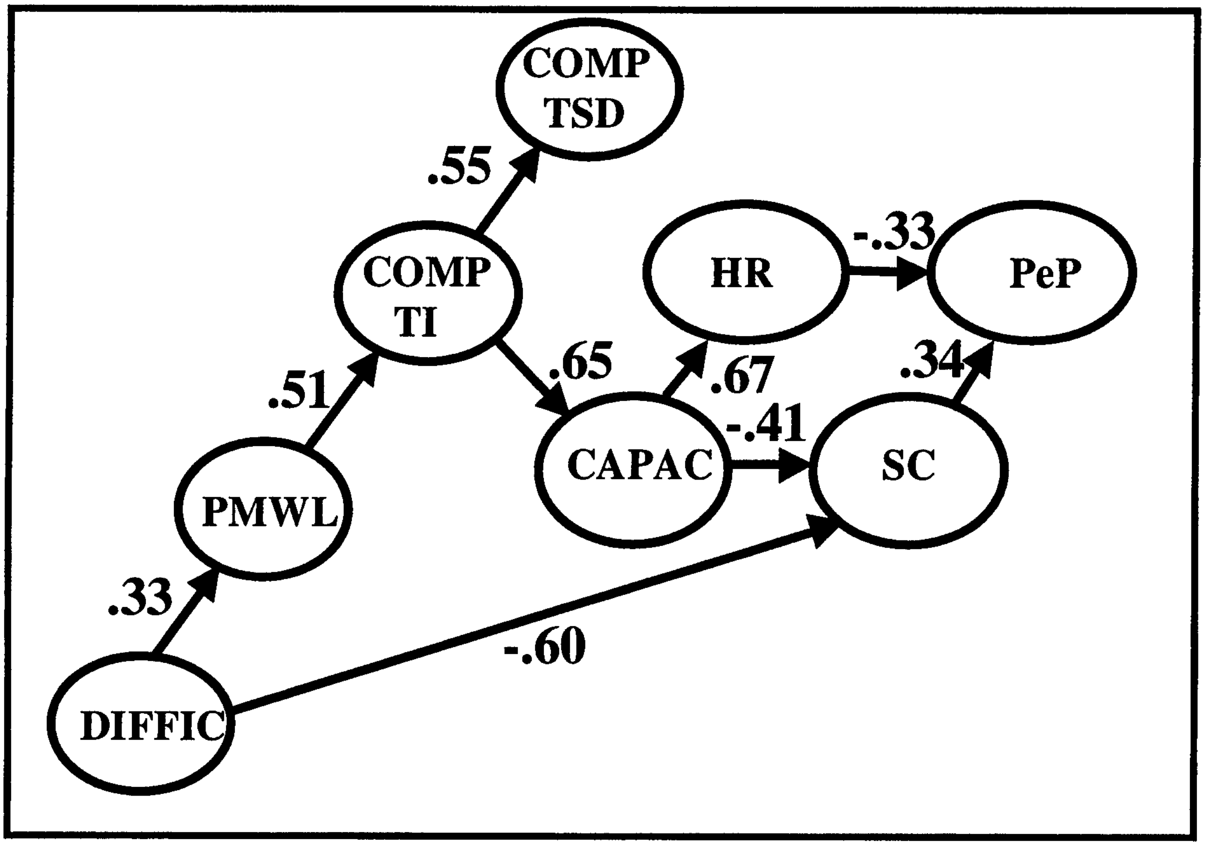

Figure 13. The final structural LISREL model of the relationships between seven of the psychological indices and the HR measure. All effects are significant ($p < .05$). Adjusted Goodness of Fit is .62 and Root Mean square is .119.

analyses[5]. This increases the random variation of the correlations and the stability of the model is reduced.

As can be seen from Figure 13, the model has its starting point in the difficulty and complexity (DIFFIC) of the missions and its terminal point in the performance of the pilots (PeP). Increasing mission difficulty is followed by an increased general mental workload (PMWL), which in turn, affects the complexity of information on TI (COMP TI) and TSD (COMP TSD). (In the model from the air, we found interaction effects between COMP TSD and COMP TI. In this model, this effect was not found in spite of the fact that the correlation between the indices was .55 ($p < .01$)).

[5] **From a statistical point of view, the number of cases is too low for a development and fair testing of a model. However, with the model from the air as a starting point and pattern of comparison the conditions are improved.**

The complexity of information on TI (COMP TI) has a strong effect on the mental capacity index (CAPAC). This means that high information complexity in the displays has a deteriorating effect on mental capacity. The markers of the mental capacity index deal with difficulties in evaluating the synthetic information and the necessity or need to reduce the flow of information. The mental capacity index (CAPAC) has a strong effect on heart rate (HR). Forty five percent of the variance in heart rate is explained by the variance in mental capacity. Thus, a decreased mental capacity results in an increased psychophysiological activation. This means that the effect of an increased information complexity on heart rate is mediated by a reduced mental capacity. Analyses of the items or markers of the capacity index show that it is the items treating the need of the pilots to reduce superfluous information that show the highest relationships with heart rate.

Mental capacity (CAPAC) has an effect on situational cognizance (SC), which in its turn, affects perceived performance (PeP). This effect was expected but not found in the model from the air. This means that a reduced mental capacity restricts the pilot's situational cognizance, which, in turn, reduces his performance. Finally, heart rate (HR) has an effect on performance (PeP). The higher the heart rate the worse the performance.

The majority of the effects found in this model are the same as those found in the model from the air (Figure 4). In the analyses of the model from the air we found that it could be divided into three consecutive parts; one consisting of aspects of mission and system demands, one comprising aspects of mental workload, information load and mental capacity, and one including situational cognizance and performance aspects. These three consecutive parts were also found in this model from the simulations. This means that there is a close correspondence between the two models. This causal sequence of system and mission demands to mental workload to situational awareness to pilot performance has come up in other studies (Svensson, 1997; Svensson et al., 1997). The way the pilot copes with the demands of the mission forms an intermediary and compensating process affecting his situational awareness and performance.

Inspection of the model from the simulations (Figure 13) shows an interesting relationship between the performance aspects perceived performance (PeP) and situational cognizance (SC) and the two workload aspects mental capacity reduction (CAPAC) and mean heart rate (HR). These four aspects were used to derive two new second order factors in order to examine the relationship between pilot mental workload and pilot performance on a more general level.

Figure 14 presents a LISREL analysis of the indices CAPAC, SC, PeP, and HR. As can be seen from the figure, the fit of the model is almost perfect. The mental capacity reduction index and heart rate were optimally combined to form a second order factor called general workload (GENWL). The situational cognizance index (SC) and the perceived performance index (PeP) were combined to form a new second order factor called perceived outcome (PO). As can be seen in the figure, the factor loadings of the second order factors are very high. And a strong negative effect of general workload on perceived outcome was found.

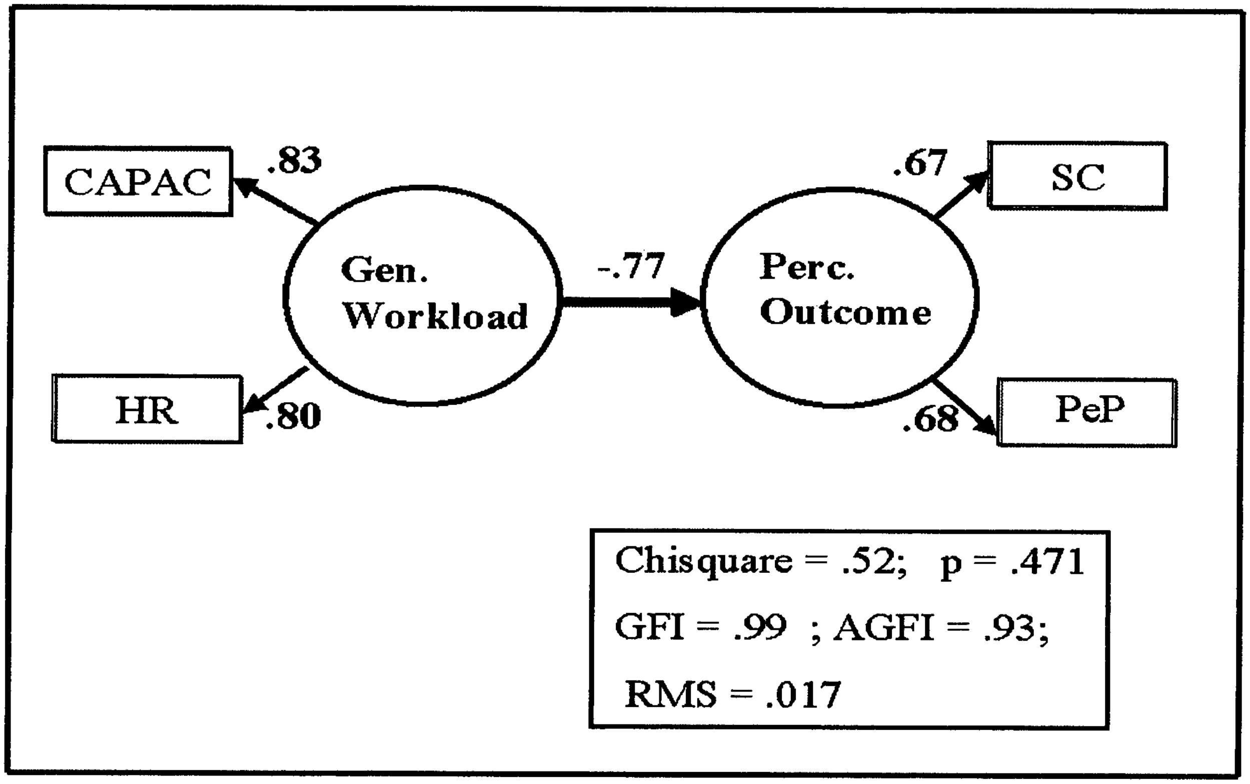

Figure 14. A structural LISREL model based on the relationships between CAPAC, HR, SC, and PeP. Adjusted Goodness of Fit is .93, and Root Mean square is .017.

The general workload factor combine psychological and psychophysiological aspects. The factor is based on seven manifest variables. During the simulation, when the pilot's mental capacity is reduced there is an increase in the heart rate.

Situational cognizance and perceived performance are equally weighted in the perceived outcome factor. The factor is based on 10 manifest variables. The perceived outcome factor shows that situational cognizance and performance are different aspects of the same concept. In previous studies, we often have considered situational awareness aspects as part of the pilot's performance (cf., Angelborg-Thanderz, 1990; Angelborg-Thanderz, 1997).

As can be seen from Figure 15, there is a curvilinear relationship between perceived outcome and general workload. In order to test for curvilinearity, separate regression analyses have been performed for workload values below and over the mean ($z=0.00$), respectively. For workload values below the mean the correlation between PO and GENWL was -.02 ($p = .93$), and for workload values over the mean it was -.59 ($p = .013$). The curvilinearity means that the performance level of the pilots is constant as long as workload is low and medium. Under higher workload levels, the performance level decreases rapidly. This empirical result is in accordance with theories of the relationship between mental workload and performance (cf.,

Lysaght et al., 1989; O'Donnell and Eggemeier, 1986). To the present authors knowledge, the theoretical relationship between the concepts has not been validated before by empirical data.

After each of the two intercepts during simulation period two, the pilots were asked to respond to questions about mental workload using the Bedford Scale (BFRS) and performance on a seven point Likert scale. Situational awareness was assessed using a scale developed within the VINTHEC-project (Svensson, Angelborg-Thanderz and van Avermaete, 1997). Thus, the pilots were asked to rate workload, situational awareness and performance directly after each intercept during the mission. The questionnaires were quick and easy to use and it took about 30 seconds to answer them.

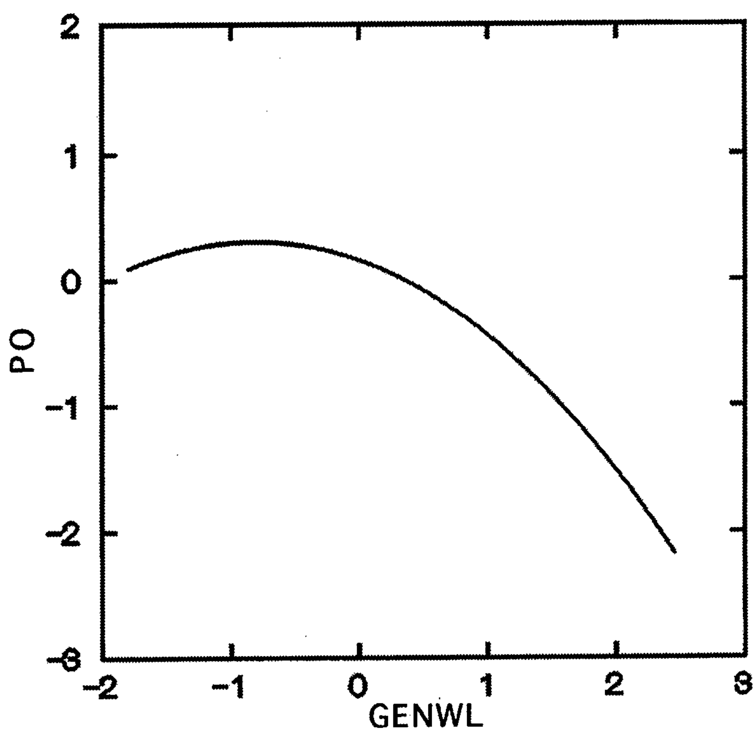

Figure 15. Perceived outcome (PO) as a function of general workload (GENWL). The scale values are standardized (z-values). The curve has been smoothed by means of distant weighted least squares regression.

In addition to mean two-minutes HR, we also used fixation rate (FR). We have used the fixation rate (FR) as a crude index of the pilots' visual search behavior. The unit of the index is the number of changes in eye point of gaze fixations per 30 seconds.

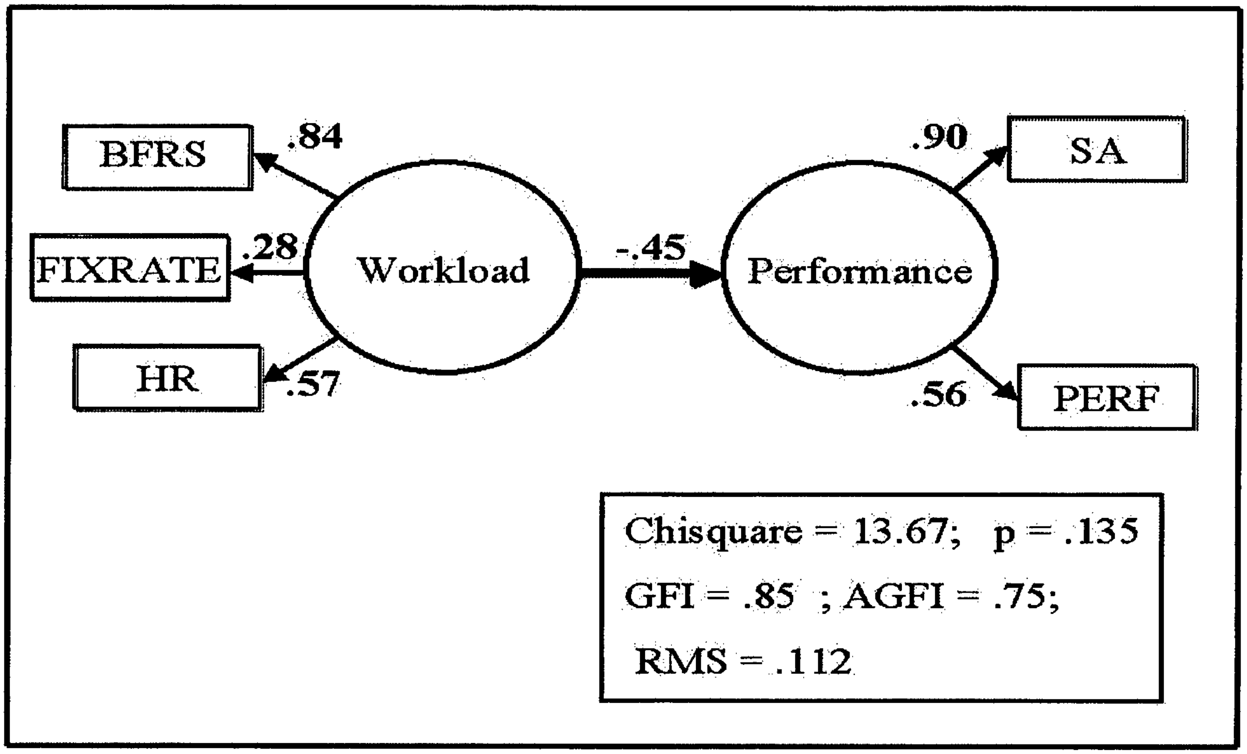

Figure 16. A structural LISREL model based on the relationships between mental workload (BFRS), fixation rate (FIXRATE), heart rate (HR), situational awareness (SA), and performance (PERF). Adjusted Goodness of Fit is .75, and Root Mean Square is .112.

We examined the ratings and psychophysiological data after each intercept. For the post intercept measures, significant correlations were found between performance ratings and the SA ratings ($r = 0.52$; $p < .01$), between mental workload and heart rate ($r = .49$; $p < .01$), between heart rate and fixation rate ($r = 0.45$; $p < .01$), and between mental workload and SA ($r = -0.46$; $p < .01$). This correlation matrix was the input for a LISREL model. The solution is shown in Figure 16.

The fit of the model is acceptable (Adjusted Goodness of Fit Index = .75 and Root Mean Square=.112). The ratings of mental workload by means of the Bedford scale (BFRS), the fixation rate (FIXRATE), and heart rate (HR) are significant markers of the workload factor. This means that an increased activity in the pilot's visual search behavior, an increase in his heart rate, and an increase in his perceived mental workload go together in a workload factor. It is of special interest that two psychophysiological variables go together with a psychological variable.

It is notable that the same structure as in Figure 14 above was derived from the LISREL procedure even though the input variables used here were different and obtained at different times. This adds credibility to the notion that the underlying structure is valid and robust.

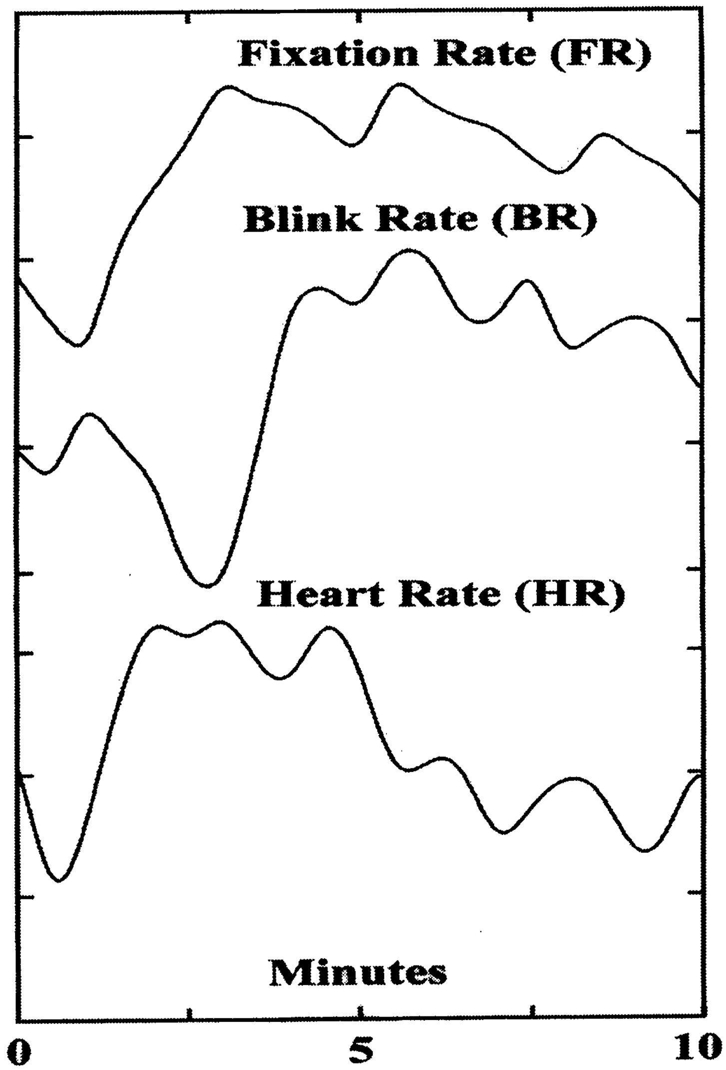

Figure 17. Data from one pilot showing the dynamic changes of fixation rate (FR), blink rate (BR), and heart rate (HR) as a function of mission time (Minutes).

Figure 17 presents an example of the dynamic changes of fixation rate (FR), blink rate (BR), and heart rate (HR) as a function of mission time. As can be seen from the figure all three measures change over time and mission phases with higher cognitive demands and higher workload (as the intercept phases) cause increases in heart rate and fixation rate, but decreases in blink rate. The first five minutes represent an intercept phase.

Visual inspection shows that this example of how fixation rate, blink rate, and heart rate change during periods of cognitive demands is representative for other pilots. And even if there are differences between the pilots with respect to the changes of the three measures our conclusion is that these interesting relationships between the measures recur during the intercepts of our scenarios.

Changes (in terms of increased fixation rate and decreased blink rate) observed prior to weapons delivery indicate that the pilots are searching the visual environment and reducing their blinking so as not to miss significant target information. Decreased blinking has the effect of reducing the probability of missing significant visual stimuli because the pilot is temporally blind during the eye closure.

At the same time heart rate was found to be increasing reaching the peak heart rate at or just after the time of weapon delivery (cf., Angelborg-Thanderz, 1990). These observations indicate that the eye and heart activity is very tied to the activities required for successful performance during the engagement. The coordination of these physiological systems is controlled by higher brain centers.

Figure 18 represents a generic model of the relationships among blink rate, fixation rate and heart rate for pilots during the air-to-air intercepts. The model or representation shows that the physiological signals seem to be coordinated in a fashion that permits optimal performance during the high demands of the intercept.

It may be possible to develop an algorithm that permits the detection of high information input (bottle necks) during intercepts prior to weapons delivery. In a first step of this development we have analyzed the relationship between fixation rate and blink rate (FR/BR) over mission time. So far we have found that this index increases during the intercepts with a maximum at

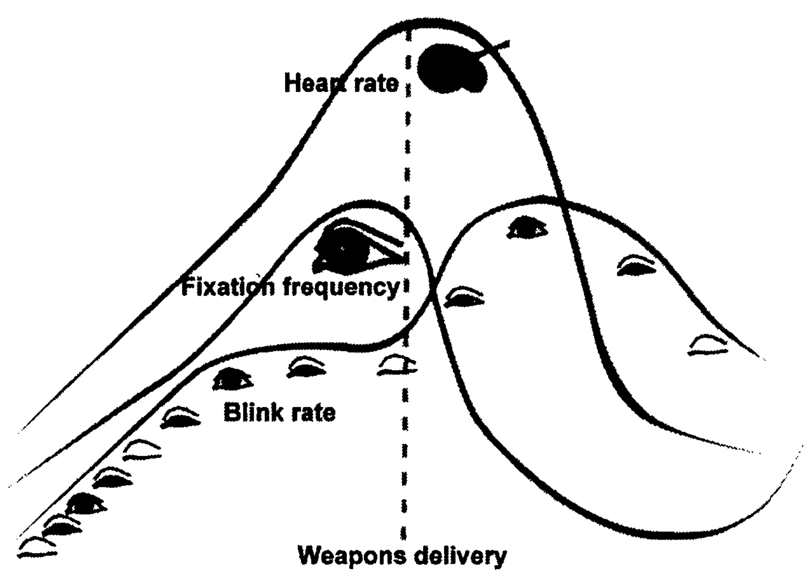

Figure 18. A generic model of the relationships among blink rate, fixation rate and heart rate for pilots during the air-to-air intercepts.

or just prior to weapons delivery. The analyses of the relationships between the blink rate, fixation rate, and heart rate will continue, and in a new study of missions in the air all three measures will be recorded.

The index could be used in training of optimal visual search behavior. It would also be useful for cockpit design and analysis so that the system will match the information processing needs of the pilot. Furthermore, data from the algorithm could be used as input to flight and weapon systems (cf., Hankins and Wilson, 1998). If the aircraft systems, by means of this technique, can adapt to the workload and performance levels of the pilot the total man machine performance will be improved.

Discussion - Simulated Missions

The same questionnaires were used in the simulation study as in the study in the air. In addition to these measures, psychophysiological variables (heart rate, blink rate, and fixation rate) were used in the simulations. Furthermore, in the second period of the simulation study, the pilots were asked to respond to questions about mental workload, performance, and situational awareness after each of two consecutive intercepts.

A wealth of empirical data shows that heart rate is a sensitive measure under different circumstances in both real and simulated missions (Wilson et al., 1987, 1988; Angelborg-Thanderz, 1990). The reactivity of heart rate to the changes in cognitive load over the mission was also documented in this study. We found, for example, significant differences in heart rate (two minute mean) between approach and intercept phases. It was also found that the 'two minutes mean' was a reliable index of psychophysiological activation and mental workload.

When comparing the mean of the fixation rate for intercept A with that of intercept B, we found a significant co-variation. This shows that the fixation rate measure is reliable and that there are inter-individual differences in the visual search behavior of the pilots. This stability of the visual search behavior of the pilots has implications for both training and cockpit design. To the present authors knowledge, this stability in visual search behavior from intercept to intercept has not been empirically validated before.

Several significant correlations were found between the psychological indices and heart rate. Of special interest is the high correlation between the mental capacity index and heart rate.

When comparing the correlation matrixes (the correlations between the indices) from the study in the air and in the simulation, we found a high concordance. This means that the internal relationships between the indices were about the same in the air and in the simulation. Accordingly, it was reasonable to use the model from the air as a starting point in the model analyses of the data from the simulations.

When comparing the model in the air with the model from the simulations, we found that the majority of the effects found are the same. In the analyses of the model from the air, we found that it could be divided into three consecutive parts; one consisting of aspects of mission and

system demands, one comprising aspects of mental workload, information load and mental capacity, and one including situational awareness and performance aspects. These three consecutive parts were also found in the model from the simulations. Accordingly, there is a close correspondence between the two models.

Inspection of the model shows an interesting relationship between the performance aspects perceived performance and situational cognizance and the two aspects of workload mental capacity reduction and mean heart rate. These four aspects were used to derive two new second order factors in order to examine more closely the relationship between pilot mental workload and pilot performance.

From the model tested, we found a general workload factor which combines psychological and psychophysiological aspects. When the pilot's mental capacity was reduced, there was an increase in heart rate. It is of interest that of the six markers of the mental capacity index, it is the items dealing with the necessity or need to reduce the flow of information that show the highest relationship with heart rate.

We also found that situational cognizance and perceived performance were equally weighted in a second order factor which we called perceived outcome. The perceived outcome factor shows that situational cognizance and performance are different aspects of the same concept. In previous studies, we often have considered situational awareness aspects as part of the pilot's performance (cf., Angelborg-Thanderz, 1990; Angelborg-Thanderz, 1997).

It is interesting to note that there is a curvilinear relationship between the general workload and perceived outcome factors. The curvilinearity means that the performance level of the pilots is constant as long as workload is low and medium. Under higher workload levels the performance level decreases rapidly. This empirical result is in accordance with theories of the relationship between mental workload and performance (cf., O'Donnell and Eggemeier, 1986; Lysaght et al., 1989). To the present authors knowledge, the theoretical relationship between the concepts has not been validated before by empirical data.

Of specific interest are the relationships among the dynamic changes in blink rate, fixation rate, and heart rate, as a function of mission time. All three measures change over time, and mission phases with higher cognitive demands and higher workload cause increases in heart rate and fixation rate, but decreases in blink rate. The changes observed prior to weapons delivery indicate that the pilots are searching the visual environment and reducing their blinking so as not to miss significant target information. Decreased blinking has the effect of reducing the probability of missing significant visual stimuli as the pilot is temporally blind during the eye closure.

At the same time heart rate was found to increase reaching the peak heart rate at or just after the time of weapon delivery. Increased heart rates have in many studies been reliably found to be correlated with increased mental activity. These observations indicate that the eye and

heart activities are very tied to the activities required for successful performance during the engagement.

CONCLUSIONS

The purposes of the studies were to (a) validate psychological, psychophysiological, and performance based measures of pilot mental workload (PMWL), situational cognizance (SC), and operative effectiveness (OE), (b) develop models of pilot performance for systems and missions evaluation, (c) compare real and simulated missions, and (d) discuss the application of these results to the systematic evaluation of systems and missions with the pilot in the loop.

Factor analyses of the data from missions in the air validated the psychological indices developed in previous studies, and we found that they have an acceptable to high reliability.

Subjective ratings made by experts have turned out to be extremely useful in many contexts. The precision of single ratings is often modest, but the grouping of ratings of different aspects into factors increases both reliability and validity. Johannsen et al. (1977) claim at "Despite all the well-known difficulties of the use of rating scales we feel that these must be regarded as central to any investigation. If the person feels loaded and effortful, he is loaded and effortful whatever the behavioral and performance measures may show."

We found that pilot mental workload is sensitive to increased information load during intercept phases of military missions. Increases in workload turn up earlier than decreases in situational cognizance and performance. These differences reflect how the pilots cope with the load of the situation. The pilots try, as long as possible, to maintain their performance and situational cognizance by increasing their mental effort. However, the performance decreases found show that this compensation does not stand firmly to the end.

The mental workload under mission phases of high complexity shows clearly that the pilots mental reserve capacity is restricted. The pilots try to shut themselves off from information, because they must focus on those aspects they consider most important. The expression 'mental tunnel vision' can summarize this condition.

We found an unfavorable relationship between mental workload, situational cognizance and performance during missions of moderate and high complexity. It is reasonable to suppose that the mission complexity and psychological stress of real war situations are higher than that of the most complex missions of this study. Thus, extrapolation of the changes in mental workload, situational cognizance, and performance found in this study indicates a sub-optimal performance of the pilots in real war situations. In previous studies, we have considered the relation of performance/workload as an efficiency measure (Angelborg-Thanderz, 1990).

From model analyses, we found that mission complexity affects different aspects of information and mental workload and that these aspects, in their turn, affect situational cognizance and pilot performance. The model tells us that there is a strong connection between

the information load on the displays and a reduced mental reserve capacity or mental overload. It is also evident from the model that both increases in general workload and information complexity decrease the situational cognizance of the pilots. That the pilots' situational cognizance grew worse as a function of high information complexity on the displays indicate a 'bottle neck' of the system.

The finding that the information complexity indices disclosed and identified the perceptual and cognitive aspects of information load is important. The relative importance of perceptual and cognitive factors can be estimated. That the workload measures were found to be more related to the cognitive aspects than to the perceptual aspects (only beta weights of the cognitive aspects are significant) support the position that cognitive aspects of information handling play a dominant role in cockpits of modern military aircraft.

The same questionnaires were used in the simulation study as in the study in the air. In addition to these measures, psychophysiological variables (heart rate, blink rate, and fixation rate) were used in the simulations.

A wealth of empirical data shows that heart rate is a sensitive measure under different circumstances in both real and simulated missions (Wilson et al., 1987, 1988; Angelborg-Thanderz, 1990). The reactivity of heart rate to the changes in cognitive load over the mission was documented also in this study. We found, for example, significant differences in heart rate between approach and intercept phases.

Several significant correlations were found between the psychological indices and heart rate. Of special interest is the high correlation between the mental capacity index and heart rate.

We also found that the blink rate measure is reliable and that there are inter-individual differences in the visual search behavior of the pilots. This stability of the pilots' visual search behavior has implications for both training and cockpit design. To the present authors knowledge this stability from intercept to intercept has not been empirically validated before.

When comparing the correlation matrixes (the correlations between the indices) from the study in the air and in the simulator, we found a high concordance. This means that the internal relationships between the indices were about the same in the air and in the simulation. Accordingly, it was reasonable to use the model from the air as a starting point in the model analyses of the data from the simulations.

When comparing the model in the air with the model from the simulations, we found that the majority of the effects found in the model from the simulator is the same as those found in the model from the air. In the analyses of the model from the air, we found that it could be divided into three consecutive parts: one consisting of aspects of missions and systems demands, one comprising aspects of mental workload, information load and mental capacity, and one including situational cognizance and performance aspects. These three consecutive parts were also found in the model from the simulations. This means that there is a close correspondence between the

two models. The high degree of similarity between the flight and simulator models adds credulance to the veracity of the models.

The physiological measures were significantly correlated with several of the subjective ratings, and both types of variables were included in the same factor structures. This strengthens the significance of the models because it included both types of measures in the same structures rather than placed them in separate structures as is often reported. This also supports the notion that mental workload is a multifaceted concept comprising both subjective and physiological aspects.

Angelborg-Thanderz (1990) has reported several relationships between simulated and real flight missions. She found a significant relation between the efficiency factor (performance/workload) in the simulation and in the real flight. About 20 percent of the variance in efficiency in the air could be explained by the variance in the efficiency in the simulations.

It was not possible to make such a strict comparison in this study. However, the similarities between the model from the air and the model from the simulations indicate that the pilots use the same mental models during real and simulated flight. It is important to remember that the models developed represent the pilots' internal representation of the relationships between the central concepts pilot mental workload, situational cognizance, and pilot performance.

Inspection of the model from the simulations shows an interesting relationship between the performance aspects perceived performance and situational cognizance and the two aspects of workload mental capacity reduction and mean heart rate. These four aspects were used to derive two new second order factors in order to examine more closely the relationship between pilot mental workload and pilot performance.

From the model tested, we found a general workload factor which combines psychological and psychophysiological aspects. When the pilot's mental capacity was reduced there was an increase in heart rate. It is of interest that of the six markers of the mental capacity index it is the items dealing with the necessity or need to reduce the flow of information that show the highest relationship with heart rate.

We also found that situational cognizance and perceived performance were equally weighted in a second order factor called perceived outcome. The perceived outcome factor shows that situational cognizance and performance are different aspects of the same concept. In former studies, we have considered situational cognizance aspects as part of the pilot's performance (cf., Angelborg-Thanderz, 1990; Angelborg-Thanderz, 1997).

It is interesting to note that there is a curve-linear relationship between the factors general workload and perceived outcome. The curve-linearity means that the performance level of the pilots is constant as long as workload is low and medium. Under higher workload levels, the performance level decreases rapidly. This empirical result is in accordance with theories of the

relationship between mental workload and performance (cf., O'Donnell and Eggemeier, 1986; Lysaght et al., 1989). To the present authors knowledge, the theoretical relationship between the concepts has not been validated before by empirical data.

Of specific interest are the relationships between the dynamic changes in blink rate, fixation rate, and heart rate as a function of mission time. All three measures change over time and mission phases with higher cognitive demands and higher workload (as the intercept phases) cause increases in heart rate and fixation rate, but decreases in blink rate. The changes observed prior to weapons delivery indicate that the pilots are searching the visual environment and reducing their blinking so as not to miss significant target information. Decreased blinking has the effect of reducing the probability of missing significant visual stimuli because the pilot is temporally blind during the eye closure.

At the same time, heart rate was found to be increasing and reaching the peak heart rate at or just after the time of weapon delivery. Increased heart rates have been found in many studies to be correlated with increased mental activity. Accordingly, these measures are sensitive to the dynamic changes of pilot mental load. These observations indicate that the eye and heart activity is tied to the activities required for successful performance during the engagement.

The combined analysis of the heart rate, blink rate and fixation rate data just prior to and during the weapons delivery yielded interesting relationships that should be pursued further. This type of detailed analysis may lead to a better understanding of the dynamics of the inter-relationships among the physiological measures that can be very useful for training and design analysis.

The results of this project may have important consequences for mission accomplishment and increased flight safety. Better models of human performance will permit the design of better systems. As aircraft become more complex, there is an increased need to incorporate the information processing characteristics of the pilot (cf., Hankins and Wilson, 1998). By means of this feedback, the systems of the aircraft can adapt to the workload and performance levels of the pilots. Otherwise, the systems will not be able to meet their goals.

Our general conclusion is that we have found and verified the internal relationships between the central aspects pilot mental workload, situational cognizance, and pilot performance. We have demonstrated how they change as a function of the complexity of the missions performed. From the model analyses, we concluded that the pilots use the same mental model during real and simulated missions. We were successful in combining psychophysiological and psychological variables into factors. This illustrates the multifacetedness of the concepts. The dynamic changes of heart rate, fixation rate, and blink rate during mission phases of high complexity show interesting relationships of importance in analyses of mental 'bottle necks'. We have also demonstrated that we, by means of reliable and valid psychological and psychophysiological measures, can analyze the interaction between the pilot and his aircraft.

REFERENCES

Angelborg-Thanderz, M. 1982, *Assessing pilot performance and mental workload in training simulators,* The Royal Aeronautical Society, London.

Angelborg-Thanderz, M. 1989, *Assessing pilot performance in training simulators. A structural analysis,* in 1989 Spring Convention - flight simulation: Assessing the benefits and economics, The Royal Aeronautical Society, London.

Angelborg-Thanderz, M. 1990, *Military flight training at a reasonable price and risk,* Economics Research Institute, Stockholm School of Economics and FOA report C 50083 - 5.1. (report in Swedish, summary in English).

Angelborg-Thanderz, M. 1997, Military pilot performance – dynamic decision making in its extreme, in *Decision Making Under Stress, Emerging Themes and Applications,* (eds.), Flin, F., Salas, E., Strub, M., and Martin, L. (Ashgate Publishing Company, Aldershot. Hants, England), 225-232.

Baddely, A. 1994, The magical number seven: still magic after all these years?, *Psychological Review, 101,* 353-356.

Berggren, P. 1998, *Situational awareness, mental workload, and pilot performance,* Master's thesis in psychology, University of Linköping, Sweden.

Boff, K.R., Kaufman, L., and Thomas, J.P. 1994, *Handbook of Perception and Human Performance.* Volume 2. Cognitive Processes and Performance.

Caldwell, J. A., Wilson, G. F., Cetinguc, M., Gallard, A. W. K., Gundel, A., Lagarde, D., Makeig, S., Myhre, G., and Wright, N. A. 1994, *Psychophysiological assessment methods,* AGARD-AR-324, Neuilly Sur Seine, France: NATO.

Carmody, M. A. 1994, *Current issues in the measurement of military aircrew performance: A consideration of the relationship between available metrics and operational concerns,* Air Vehicle and Crew Systems Technology Department, Naval Air Warfare Center: Aircraft Division, Warminster, PA.

Casali, J.G., and Wierwille, W.W. 1983, A comparison of rating scale, secondary task, physiological, and primary task workload estimation techniques in a simulated flight task emphasizing communication load, *Human Factors, 25,* 623-642.

Chi, M.T.H., Glaser, R., and Farr, M.J. (Eds) 1988, *The nature of expertise* (Hillsdale, New Jersey: Lawrence Erlbaum).

Cooper, G.E., Harper, R., Jr. 1969, *The use of pilot rating in the evaluation of aircraft handling qualities,* NASA, Moffett Field, CA, Report No. NASA TN-D-5153.

Cronbach, L. J. 1966, *Essentials of psychological testing* (Harper and Brothers, Publishers, New York).

Easterbrook, J.A. 1959, The effect of emotion on cue utilization and the organization of behavior, *Psychological Review, 66,* 183-201.

Eggemeier, F. T. 1984, *Workload metrics for system evaluation,* in proceedings of the NATO Defence Research Group Panel VIII Workshop on 'Applications of Systems Ergonomics to Weapon System Development', Schrivenham, UK: Royal Military College of Science.

Eggemeier, F. T., Biers, D. W., Wickens, C. D., Andre, A. D., Vreuls, D., Billman, E. R., and Schueren, J. 1990, *Performance assessment and workload evaluation systems: Analysis of candidate measures,* Technical Report No. HSD-TR-90-023, Brooks Air Force Base, TX: Armstrong Aerospace Medical Research Laboratory.

Eggemeier, F.T., and Wilson, G.F. 1991, Performance-based and subjective assessment of workload in multi-task environments, in D.L. Damos (ed) *Multiple task performance* (Taylor and Francis Ltd, London), 217-278.

Endsley, M. R. 1988, *Design and evaluation for situation awareness,* in Proceedings of the Human Factors Society 32th Annual Meeting, Santa Monica, CA, Human Factors Society.

Endsley, M. R. 1994, Situation awareness in dynamic human decision making: Measurement, in R. D. Gilson, D. Garland, and J. M. Koonce (eds) *Situational awareness in complex systems* (Embry-Riddle Aeronautical University Press, Daytona Beach).

Endlsey, M. R. 1995, Theoretical underpinnings of situational awareness: A critical review, in D. Garland, and M. R. Endsley (eds) *Experimental Analysis and Measurement of Situational Awareness* (Embry-Riddle Aeronautical University Press, Daytona Beach).

Farmer, E. 1993, *Conceptual issues in workload,* in Workload assessment and aviation safety, The Royal Aeronautical Society, London.

Gawron, V.J., Schiflett, S.G., and Miller, J.C. 1989, Measures of in-flight workload, in R.S. Jensen (ed) *Aviation psychology* (Gower Publishing Company, Aldershot), 240-287.

Gilson, R.D. (special issue editor). 1995, *Situational awareness* (special issue), *Human Factors, 37.*

Gopher, D., and Donchin E. 1986, Workload - an examination of the concept, in K.R. Boff, L. Kaufman and J.P. Thomas (eds) *Handbook of perception and human performance. Vol.II.* (John Wiley and Sons, New York), 41-1-41-49.

Hancock, P. A., and Meshkati, N. 1988, *Human mental workload* (Elsevier, Amsterdam).

Hankins, T.C., and Wilson, G.F. 1998, A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight, *Aviation, Space, and Environmental Medicine, 69,* 360-367.

Harris, R.L., Sr., and Christhilf, D. M. 1980, *What do pilots see in displays?*, in Proceedings of the Human Factors Society Meeting, Santa Monica, CA: Human Factors Society.

Hart, S. G., and Staveland, L. E. 1988, Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, in P.A. Hancock and N. Meshkati (eds) *Human Mental Workload* (Elsevier, Amsterdam), 139-183.

Hart, S.G., and Wickens, C.D. 1990, Workload assessment and prediction, in H.R. Booher (ed) *MANPRINT. An approach to systems integration* (van Nostrand Reinhold, New York), 257-296.

Itoh, Y., Hayashi, Y., Tsukui, I., and Saito, S. 1990, The ergonomic evaluation of eye movements and mental workload in aircraft pilots, *Ergonomics, 33,* 719-733.

Jensen, S.E. 1994, Developing a flight workload profile using Continuous Subjective Assessment of Workload (C-SAW), in R. Fuller, N. Johnston and N. McDonald (eds) *Human factors in aviation operations* (Averbury Aviation, Aldershot), 307-312.

Johannsen, G., Moray, N., Pew, R., Rasmussen, J., Sanders, A., and Wickens, C. 1977, Final report of experimental psychology group, in N. Moray (ed) *Mental workload: Its theory and measurement* (Plenum press, New York).

Jöreskog K.G., and Sörbom, D. 1984, *LISREL VI. Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods,* Department of Statistics, University of Uppsala, Sweden.

Kahneman, D. 1973, *Attention and effort,* (Englewood Cliffs, New Jersey: Prentice-Hall).

Kantowitz, B. H. 1993, *Pilot workload and flight deck automation,* in Workload assessment and aviation safety, The Royal Aeronautical Society, London.

Katoh, Z., Kadoo, A., Itoh, H., and Maruta, H. 1995, Eye scanning behavior as a discriminator of instrument- and scenery-centered flight task, *Journal of Human Ergonomy, 24,* 80-84.

Kennedy, R. S., Braun, C. C., and Massey, C. J. 1995, *Performance, workload, and eye movement activity,* Report No. SAE Paper 951570, SAE, 25th International Conference on Environmental Systems, San Diego, CA.

Klein, G.A. 1993, *Naturalistic decision making: implications for design,* Crew Systems Ergonomics, Information Analysis Center, Wright-Patterson AFB, Report No. CSERIAC SOAR 93-1.

Lysaght, R.J., Hill, S.G., Dick, A.O., Plamondon, B.D., Linton, P.M., Wierwille, W.W., Zaklad, A.L., Bittner, Jr., A.C., and Wherry, R.J. 1989, *Operator workload: Comprehensive review and evaluation of operator workload methodologies*, United States Army Research Institute for the Behavioral Sciences, Technical Report 851, MDA 903-86-C-0384.

May, J. G., Kennedy, R. S., Williams, M. C., Dunlap, W. P., and Brannan, J. R. 1990, Eye movements indices of mental workload, *Acta Psychologica, 75,* 75-89.

McMillan, G. R., Bushman, J. and C. L. A. Judge. 1996, *Keynote Address: - Evaluating Pilot Situational Awareness in an Operational Environment*, in AGARD Conference Proceedings 575, Situational Awareness: Limitations and Enhancement in the Aviation Environment, Brussels.

Meister, D., and Hogg, D. 1994, Cognitive error in complex, uncertain, and risky work situations, in Aghazadeh, F. (ed*) Advances in Industrial Ergonomics and Safety VI* (Taylor and Frances).

Miller, G.A. 1956, The magical number seven, plus or minus two, *Psychological Re-view, 63,* 81-97.

Moray, N. (ed). 1979, *Mental workload: Its Theory and Measurement* (Plenum, New York).

Muckler, F.A., and Seven, S.A. 1992, Selecting performance measures: "Objective" versus "subjective" measurement, *Human Factors, 34,* 441-55.

Musashi, M. 1645, *A book of five rings,* Translation into English by Harris, V. London: Fontana.

O'Donnell, R.D., and Eggemeier, F.T. 1986, Workload assessment methodology, in K.R. Boff, L. Kaufman and J.P. Thomas (eds) *Handbook of perception and human performance. Vol. II* (John Wiley and Sons, New York), 42-1 - 42-49.

van Raaij, W.F. 1988, Information processing and decision making. Cognitive aspects of economic behavior, in W. F. van Raaij, G. M. van Veldhoven and K. E. Wärneryd (eds) *Handbook of Economic Psychology* (Kluwer Academic Publishers, Dordrecht), 74-106.

Reid, G.B., and Nygren, T.E. 1988, The subjective workload assessment technique: A scaling procedure for measuring mental workload, in P.A. Hancock and N. Meshkati (eds) *Human mental workload* (Elsevier, Amsterdam), 184-218.

Roscoe, A. H. 1987, In-flight assessment of workload using pilot ratings and heart rate, in A.H. Roscoe (ed) *The practical assessment of pilot workload*, AGARDograph No, 282.

Roscoe, A. H. 1992, Assessing pilot workload. Why measure heart rate, HRV and respiration?, *Biological Psychology, 34*, 259-287.

Roscoe, A. H., and Ellis, G. A. 1990, *A subjective rating scale for assessing pilot workload in flight: A decade of practical use,* Royal Aerospace Establishment, Technical Report 90019.

Schiffrin, S. S., Reynolds, M. L., and Young, F. W. (1981). *Introduction to multi-dimensional scaling theory, Methods, and applications*, Academic Press, New York.

Simon, H.A. 1990, Invariants of human behavior, *Annual Review of Psychology, 41,* 1-19.

Sjöberg, L., Svensson, E., and Persson, L.O. 1979, The measurement of mood, *Scandinavian Journal of Psychology, 20,* 1-18.

Stern, J. A., Boyer, D., Schroeder, D., Touchstone, M., and Stoliarov, N. 1994, *Blinks, saccades, and fixation pauses during vigilance task performance: Time on task,* Report No. DOT/FFA/AM-94/26, Washington University, Seattle.

Svenson, O., and Benson, L. 1991, *Framing and time pressure in decision making,* Psychological Research Bulletin XXXI:1, Department of Psychology, University of Lund, Sweden.

Svenson, O., and Edland, A. 1987. Change of preferences under time pressure: choices and judgments, *Scandinavian Journal of Psychology, 28,* 322-330.

Svensson, E., 1997, Pilot mental workload and situational awareness – psychological models of the pilot, in *Decision Making Under Stress, Emerging Themes and Applications*, (eds.), Flin, F., Salas, E., Strub, M. and Martin, L. (Ashgate Publishing Company, Aldershot. Hants, England), 261-267.

Svensson, E., Angelborg-Thanderz M., Sjöberg, L., and Gillberg, M. 1988, Military flight experience and sympatho-adrenal activity, *Aviation, Space, and Environmental Medicine, 59,* 411-416.

Svensson, E., Angelborg-Thanderz, M., and Sjöberg, L. 1991, *Mental workload in military aviation,* The 27th International Applied Military Psychology Symposium, Stockholm: Department of Defence, USA, Naval Research, European Office (ONREUR), and National Defence Research Establishment, Sweden.

Svensson, E., Angelborg-Thanderz, M., and Sjöberg, L. 1993a, Mission challenge, mental workload and performance in military aviation, *Aviation, Space, and Environmental Medicine, 64,* 985-991.

Svensson, E., Angelborg-Thanderz, M., and Sjöberg, L. 1993b, *Information overflow? Mental workload and performance in combat aircraft,* In Workload Assessment and Aviation Safety. London: The Royal Aeronautical Society.

Svensson, E., and Angelborg-Thanderz, M. 1994, *Systems evaluation - Mental workload and Performance.* Presentation (E. Svensson) at the DIF-seminar "Samspillet Menneske – Maskine,"Köpenhamn, Danmark. (Abstract.)

Svensson, E., and Angelborg-Thanderz, M. 1995, Mental workload and performance in combat aircraft: systems evaluation, In Fuller, R., Johnston, N., and McDonald, N. (Eds), *Human Factors in Aviation Operations* (Aldershot, Hants, England: Avebury aviation).

Svensson, E., Angelborg-Thanderz, M., and van Awermaete, J. 1997, *Dynamic measures of pilot mental workload, pilot performance, and situational awareness.* Technical Report: VINTHEC-WP3-TR01. NLR, Amsterdam.

Svensson, E., Angelborg-Thanderz, M., Sjöberg, L., and Olsson, S. 1997, Information complexity - mental workload and performance in combat aircraft, *Ergonomics, 40,* 362-380.

Thanderz, M. 1973, *A study of pilot training on a modern weapon system,* Institute of Aviation Medicine, Flight Test Centre, Air Materiel Department, Defence Materiel Administration, Sweden, Report FMI-73-10.

Thanderz, M. 1982, *Assessing pilot performance and mental workload in training simulators,* Flight simulation, aviation medicine, and avionics system groups two day international conference, The Royal Aeronautical Society, London.

Tsang, P., and Wilson, G. F., 1997, Mental workload, in Salvendy, G., (ed.), *Handbook of Human Factors and Ergonomics,* (Wiley and Sons, New York), 417-449.

Vreuls, D., and Obermayer, R.W. 1985, Human-system performance measurement in training simulators, *Human Factors, 27,* 241-250.

Wierwille, W.W. 1979, Physiological measures of aircrew mental workload, *Human Factors, 21,* 575-593.

Wierwille, W.W., and Casali, J.G. 1983, A validated rating scale for global mental workload measurement applications, *Proceedings of the Human Factors Society, 27,* 129-133.

Wierwille, W.W., and Connor, S.A. 1983, Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator, *Human Factors, 25,* 1-16.

Williges, R.C., and Wierwille W.W. 1979, Behavioral measures of aircrew mental workload, *Human Factors, 21,* 549-574.

Wilson, G. F. 1991, *Progress in the psychophysiological assessment of workload,* Technical Report No. AL-TR-1992-0007, Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

Wilson, G. F., 1992, (ed), Cardiorespiratory measures and their role in studies of performance; Special Issue, *Biological Psychology,* 34, 91-290.

Wilson, G. F., 1993, Air-to ground training missions: a psychophysiological workload analysis, *Ergonomics, 36*: 1071-1087.

Wilson, G.F., and Eggemeier, F.T. 1991, Psychophysiological assessment of workload in multi-task environments, in D.L. Damos (ed) *Multiple task performance* (Taylor and Francis Ltd, London), 329-360.

Wilson, G. F., and Fisher, F. 1991, The use of cardiac and eye blink measures to determine flight segments in F4 crews, *Aviation, Space and Environmental Medicine, 62*:959-962.

Wilson, G. F., and Fullenkamp, P. 1991, *A comparison of pilot and WSO workload during training missions using psychophysiological data,* in E. Farmer (ed) Proceedings of the Western European Association for Aviation Psychology (Avebury, Aldershot).

Wilson, G. F., O'Donnell, R. D., and Wilson, L., 1983, *Neurophysiological measures of A-10 workload during simulated low altitude missions,* AFAMLR Technical Report No. 83-0003. Wright-Patterson AFB: Air Force Aerospace Medical Research Laboratory.

Wilson, G. F., Purvis, B., Skelly, J., Fullenkamp, P., and Davis, I. 1987, *Physiological data used to measure pilot workload in actual flight and simulator conditions,* in Proceedings of the Human Factors Society - 31st Annual Meeting.

Wilson, G. F., Skelly, J., and Purvis, B. 1988, *Reactions to emergency situations in actual and simulated flight,* in AGARD AMP Symposium on "Human behavior in high stress situations in aerospace operations," the Hague, 25-28 Oct.